

# Goals, Motivation for, and Outcomes of Personal Learning through Networks: Results of a Tweetstorm

Citation for published version (APA):

Sie, R., Patariaia, N., Boursinou, E., Rajagopal, K., Margaryan, A., Falconer, I., Bitter-Rijkema, M., Littlejohn, A., & Sloep, P. (2013). Goals, Motivation for, and Outcomes of Personal Learning through Networks: Results of a Tweetstorm. *Educational Technology & Society*, 16(3), 59-75.

## Document status and date:

Published: 01/07/2013

## Document Version:

Publisher's PDF, also known as Version of record

## Document license:

CC BY-NC-ND

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

<https://www.ou.nl/taverne-agreement>

## Take down policy

If you believe that this document breaches copyright please contact us at:

[pure-support@ou.nl](mailto:pure-support@ou.nl)

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 05 May. 2023

Open Universiteit  
[www.ou.nl](http://www.ou.nl)



*Journal of  
Educational Technology  
& Society*

*Published by:  
International Forum of Educational Technology & Society*

**<http://www.ifets.info>**

*ISSN: 1436-4522 (Online)  
ISSN: 1176-3647 (Print)*

**July 2013  
Volume 16 Issue 3**

# Educational Technology & Society

An International Journal

## Aims and Scope

*Educational Technology & Society* is a quarterly journal published in January, April, July and October. *Educational Technology & Society* seeks academic articles on the issues affecting the developers of educational systems and educators who implement and manage such systems. The articles should discuss the perspectives of both communities and their relation to each other:

- Educators aim to use technology to enhance individual learning as well as to achieve widespread education and expect the technology to blend with their individual approach to instruction. However, most educators are not fully aware of the benefits that may be obtained by proactively harnessing the available technologies and how they might be able to influence further developments through systematic feedback and suggestions.

- Educational system developers and artificial intelligence (AI) researchers are sometimes unaware of the needs and requirements of typical teachers, with a possible exception of those in the computer science domain. In transferring the notion of a 'user' from the human-computer interaction studies and assigning it to the 'student', the educator's role as the 'implementer/ manager/ user' of the technology has been forgotten.

The aim of the journal is to help them better understand each other's role in the overall process of education and how they may support each other. The articles should be original, unpublished, and not in consideration for publication elsewhere at the time of submission to *Educational Technology & Society* and three months thereafter.

The scope of the journal is broad. Following list of topics is considered to be within the scope of the journal:

Architectures for Educational Technology Systems, Computer-Mediated Communication, Cooperative/ Collaborative Learning and Environments, Cultural Issues in Educational System development, Didactic/ Pedagogical Issues and Teaching/Learning Strategies, Distance Education/Learning, Distance Learning Systems, Distributed Learning Environments, Educational Multimedia, Evaluation, Human-Computer Interface (HCI) Issues, Hypermedia Systems/ Applications, Intelligent Learning/ Tutoring Environments, Interactive Learning Environments, Learning by Doing, Methodologies for Development of Educational Technology Systems, Multimedia Systems/ Applications, Network-Based Learning Environments, Online Education, Simulations for Learning, Web Based Instruction/ Training

## Editors

**Kinshuk**, Athabasca University, Canada; **Demetrios G Sampson**, University of Piraeus & ITI-CERTH, Greece; **Nian-Shing Chen**, National Sun Yat-sen University, Taiwan.

## Editors' Advisors

**Ashok Patel**, CAL Research & Software Engineering Centre, UK; **Reinhard Oppermann**, Fraunhofer Institut Angewandte Informationstechnik, Germany

## Editorial Assistant

**Barbara Adamski**, Athabasca University, Canada; **Natalia Spyropoulou**, University of Piraeus & Centre for Research and Technology Hellas, Greece

## Associate editors

**Vladimir A Fomichov**, K. E. Tsiolkovsky Russian State Tech Univ, Russia; **Olga S Fomichova**, Studio "Culture, Ecology, and Foreign Languages", Russia; **Piet Kommers**, University of Twente, The Netherlands; **Chul-Hwan Lee**, Incheon National University of Education, Korea; **Brent Muirhead**, University of Phoenix Online, USA; **Erkki Sutinen**, University of Joensuu, Finland; **Vladimir Uskov**, Bradley University, USA.

## Assistant Editors

**Yuan-Hsuan (Karen) Lee**, National Chiao Tung University, Taiwan; **Wei-Chieh Fang**, National Sun Yat-sen University, Taiwan.

## Advisory board

**Ignacio Aedo**, Universidad Carlos III de Madrid, Spain; **Mohamed Ally**, Athabasca University, Canada; **Luis Anido-Rifon**, University of Vigo, Spain; **Gautam Biswas**, Vanderbilt University, USA; **Rosa Maria Bottino**, Consiglio Nazionale delle Ricerche, Italy; **Mark Bullen**, University of British Columbia, Canada; **Tak-Wai Chan**, National Central University, Taiwan; **Kuo-En Chang**, National Taiwan Normal University, Taiwan; **Ni Chang**, Indiana University South Bend, USA; **Yam San Chee**, Nanyang Technological University, Singapore; **Sherry Chen**, Brunel University, UK; **Bridget Cooper**, University of Sunderland, UK; **Darina Dicheva**, Winston-Salem State University, USA; **Jon Dron**, Athabasca University, Canada; **Michael Eisenberg**, University of Colorado, Boulder, USA; **Robert Farrell**, IBM Research, USA; **Brian Garner**, Deakin University, Australia; **Tiong Goh**, Victoria University of Wellington, New Zealand; **Mark D. Gross**, Carnegie Mellon University, USA; **Roger Hartley**, Leeds University, UK; **J R Isaac**, National Institute of Information Technology, India; **Mohamed Jemni**, University of Tunis, Tunisia; **Mike Joy**, University of Warwick, United Kingdom; **Athanasios Karoulis**, Hellenic Open University, Greece; **Paul Kirschner**, Open University of the Netherlands, The Netherlands; **William Klemm**, Texas A&M University, USA; **Rob Koper**, Open University of the Netherlands, The Netherlands; **Jimmy Ho Man Lee**, The Chinese University of Hong Kong, Hong Kong; **Ruddy Lelouche**, Universite Laval, Canada; **Tzu-Chien Liu**, National Central University, Taiwan; **Rory McGreal**, Athabasca University, Canada; **David Merrill**, Brigham Young University - Hawaii, USA; **Marcelo Milrad**, Växjö University, Sweden; **Riichiro Mizoguchi**, Osaka University, Japan; **Permanand Mohan**, The University of the West Indies, Trinidad and Tobago; **Kiyoshi Nakabayashi**, National Institute of Multimedia Education, Japan; **Hiroaki Ogata**, Tokushima University, Japan; **Toshio Okamoto**, The University of Electro-Communications, Japan; **Jose A. Pino**, University of Chile, Chile; **Thomas C. Reeves**, The University of Georgia, USA; **Norbert M. Seel**, Albert-Ludwigs-University of Freiburg, Germany; **Timothy K. Shih**, Tamkang University, Taiwan; **Yoshiaki Shindo**, Nippon Institute of Technology, Japan; **Kevin Singley**, IBM Research, USA; **J. Michael Spector**, Florida State University, USA; **Slavi Stoyanov**, Open University, The Netherlands; **Timothy Teo**, Nanyang Technological University, Singapore; **Chin-Chung Tsai**, National Taiwan University of Science and Technology, Taiwan; **Jie Chi Yang**, National Central University, Taiwan; **Stephen J.H. Yang**, National Central University, Taiwan.

## Executive peer-reviewers

<http://www.ifets.info/>

## Supporting Organizations

Centre for Research and Technology Hellas, Greece  
Athabasca University, Canada

## Subscription Prices and Ordering Information

For subscription information, please contact the editors at [kinshuk@ieee.org](mailto:kinshuk@ieee.org).

## Advertisements

*Educational Technology & Society* accepts advertisement of products and services of direct interest and usefulness to the readers of the journal, those involved in education and educational technology. Contact the editors at [kinshuk@ieee.org](mailto:kinshuk@ieee.org).

## Abstracting and Indexing

*Educational Technology & Society* is abstracted/indexed in Social Science Citation Index, Current Contents/Social & Behavioral Sciences, ISI Alerting Services, Social Scisearch, ACM Guide to Computing Literature, Australian DEST Register of Refereed Journals, Computing Reviews, DBLP, Educational Administration Abstracts, Educational Research Abstracts, Educational Technology Abstracts, Elsevier Bibliographic Databases, ERIC, Inspec, Technical Education & Training Abstracts, and VOCED.

## Guidelines for authors

Submissions are invited in the following categories:

- Peer reviewed publications: Full length articles (4000 - 7000 words)
- Book reviews
- Software reviews
- Website reviews

All peer review publications will be refereed in double-blind review process by at least two international reviewers with expertise in the relevant subject area. Book, Software and Website Reviews will not be reviewed, but the editors reserve the right to refuse or edit review.

For detailed information on how to format your submissions, please see:  
<http://www.ifets.info/guide.php>

## Submission procedure

Authors, submitting articles for a particular special issue, should send their submissions directly to the appropriate Guest Editor. Guest Editors will advise the authors regarding submission procedure for the final version.

All submissions should be in electronic form. The editors will acknowledge the receipt of submission as soon as possible.

The preferred formats for submission are Word document and RTF, but editors will try their best for other formats too. For figures, GIF and JPEG (JPG) are the preferred formats. **Authors must supply separate figures** in one of these formats besides embedding in text.

Please provide following details with each submission: ■ Author(s) full name(s) including title(s), ■ Name of corresponding author, ■ Job title(s), ■ Organisation(s), ■ Full contact details of ALL authors including email address, postal address, telephone and fax numbers.

The submissions should be uploaded at [http://www.ifets.info/ets\\_journal/upload.php](http://www.ifets.info/ets_journal/upload.php). In case of difficulties, please contact [kinshuk@ieee.org](mailto:kinshuk@ieee.org) (Subject: Submission for Educational Technology & Society journal).



### Full length articles

Effects of Thinking Style and Spatial Ability on Anchoring Behavior in Geographic Information Systems <i>Dai-Yi Wang, Mei-Hsuan Lee and Chuen-Tsai Sun</i>	1–13
Am I Extravert or Introvert? Considering the Personality Effect Toward e-Learning System <i>Amal Al-Dujaily, Jieun Kim and Hokyoung Ryu</i>	14–27
Leveraging Quiz-based Multiple-prize Web Tournaments for Reinforcing Routine Mathematical Skills <i>Ana I. González-Tablas, José M. de Fuentes, Jorge L. Hernández-Ardieta and Benjamín Ramos</i>	28–43
Assessment of Animated Self-Directed Learning Activities Modules for Children's Number Sense Development <i>Der-Ching Yang and Mao-Neng Li</i>	44–58
Goals, Motivation for, and Outcomes of Personal Learning through Networks: Results of a Tweetstorm <i>Rory L. L. Sie, Nino Pataraia, Eleni Boursinou, Kamakshi Rajagopal, Anoush Margaryan, Isobel Falconer, Marlies Bitter-Rijkema, Allison Littlejohn and Peter B. Sloep</i>	59–75
A Comparative Evaluation of E-learning and Traditional Pedagogical Process Elements <i>Damjan Vavpotič, Boštjan Žvanut and Irena Trobec</i>	76–87
A Novel Method for Learner Assessment Based on Learner Annotations <i>Fakhroddin Noorbehhani, Elaheh Biglar Beigi Samani and Hossein Hadian Jazi</i>	88–101
Fostering 5th Grade Students' Understanding of Science via Salience Analogical Reasoning in On-line and Classroom Learning Environments <i>Ming-Hua Chuang and Hsiao-Ching She</i>	102–118
The Increasing Acceptance of Onscreen Marking – the 'Tablet Computer' Effect <i>David Coniam</i>	119–129
Blog-Assisted Learning in the ESL Writing Classroom: A Phenomenological Analysis <i>Ming Huei Lin, Nicholas Groom and Chin-Ying Lin</i>	130–139
Prompts-based Scaffolding for Online Inquiry: Design Intentions and Classroom Realities <i>Meilan Zhang</i>	140–151
The Design of Social Agents That Introduce Self-reflection in a Simulation Environment <i>Pei-Lan Lei, Sunny S. J. Lin, Dai-Yi Wang and Chuen-Tsai Sun</i>	152–166
An Evaluation of the Learning Effectiveness of Concept Map-Based Science Book Reading via Mobile Devices <i>Che-Ching Yang, Gwo-Jen Hwang, Chun-Ming Hung and Shian-Shyong Tseng</i>	167–178
On the Educational Validity of Research in Educational Technology <i>Chris Tompsett</i>	179–190
A Review of Intervention Studies On Technology-assisted Instruction From 2005-2010 <i>Ying-Tien Wu, Huei-Tse Hou, Fu-Kwun Hwang, Min-Hsien Lee, Chih-Hung Lai, Guo-Li Chiou, Silvia Wen-Yu Lee, Yu-Chen Hsu, Jyh-Chong Liang, Nian-Shing Chen and Chin-Chung Tsai</i>	191–203
Development of a Virtual Second Life Curriculum Using Currere Model <i>Chih-Feng Chien, Trina Davis, Patrick Slattery, Wendy Keeney-Kennicutt and Janet Hammer</i>	204–219
Remote Laboratories: Design of Experiments and Their Web Implementation <i>Ananda Maiti and Balakrushna Tripathy</i>	220–233
Effects of Segmented-Animation in Projected Presentation Condition <i>Ahmad Zamzuri Mohamad Ali</i>	234–245

The Framework of Viewing and Representing Skills Through Digital Text <i>Kay Yong, Khoo I and Daniel Churchill</i>	246–258
Computer Enhanced English Language Tool for Students with Hearing Loss – A Bulgarian study <i>Milen Zamfirov and Svetoslava Saeva</i>	259–273
Effects of Touch Technology-based Concept Mapping on Students' Learning Attitudes and Perceptions <i>Gwo-Jen Hwang, Chih-Hsiang Wu and Fan-Ray Kuo</i>	274–285
Learning by Searching: A Learning Environment that Provides Searching and Analysis Facilities for Supporting Trend Analysis Activities <i>Chengjiu Yin, Han-Yu Sung, Gwo-Jen Hwang, Sachio Hirokawa, Hui-Chun Chu, Brendan Flanagan, Yoshiyuki Tabata</i>	286–300
Scaffolding Environment for e-Learning through Cloud Computing <i>Marijana Despotović-Zrakić, Konstantin Simić, Aleksandra Labus, Aleksandar Milić and Branislav Jovanić</i>	301–314
Introducing 1 to 1 in the Classroom: A Large-scale Experience in Chile <i>Magdalena Claro, Miguel Nussbaum, Ximena López and Anita Díaz</i>	315–328
Analysis of Students' Performances during Lab Sessions of Computer Networks Course <i>Bulent Gursel Emiroglu and Seda Sahin</i>	329–346
Classroom-based Cognitive Diagnostic Model For a Teacher-made Fraction- Decimal Test <i>Tsai-Wei Huang and Pei-Chen Wu</i>	347–361
Context-aware and Personalization Method in Ubiquitous Learning Log System <i>Mengmeng Li, Hiroaki Ogata, Bin Hou, Noriko Uosaki and Kousuke Mouri</i>	362–373
Learning through Multi-touch Interfaces in Museum Exhibits: An Empirical Investigation <i>Panagiotis Zaharias, Despina Michael and Yiorgos Chrysanthou</i>	374–384
How University Students Evaluate Online Information about a Socio-scientific Issue and the Relationship with their Epistemic Beliefs <i>Fang-Ying Yang, Yu-Hsin Chen and Meng-Jung Tsai</i>	385–399
The Effect of an Example-Based Dynamic Program Visualization Environment on Students' Programming Skills <i>Mehmet Tekdal</i>	400–410
Computer Based Assessment Acceptance: A Cross-cultural Study in Greece and Mexico <i>Vasileios Terzis, Christos N. Moridis, Anastasios A. Economides and Genaro Rebolledo Mendez</i>	411–424

## Book Reviews

Perspectives on Open and Distance Learning: Open Educational Resources: An Asian Perspective <i>Reviewer: Krista Jensen</i>	425–427
--	---------

## Effects of Thinking Style and Spatial Ability on Anchoring Behavior in Geographic Information Systems

Dai-Yi Wang<sup>1\*</sup>, Mei-Hsuan Lee<sup>2</sup> and Chuen-Tsai Sun<sup>3</sup>

<sup>1</sup>Department of Computer Science and Communication Engineering, Providence University, Taichung City, Taiwan

<sup>2</sup>HsinChu Chien Kung Senior High School, HsinChu City, Taiwan // <sup>3</sup>Department of Computer Science, National Chiao Tung University, HsinChu City, Taiwan // ophelia.wang@gmail.com // lmh@mail.ckjh.hc.edu.tw // ctsun@cs.nctu.edu.tw

\*Corresponding author

(Submitted December 02, 2011; Revised May 10, 2012; Accepted July 11, 2012)

### ABSTRACT

The authors propose an instructional use for Google Earth (a GIS application) as an anchoring tool for knowledge integration. Google Earth can be used to support student explorations of world geography based on Wikipedia articles on earth science and history topics. We asked 66 Taiwanese high-school freshmen to make place marks with explanatory notes and summaries (known as anchors) to serve as geographic references; they used anchors that were predefined by their teacher for subsequent geographic searches. Our investigation focused on the processes used to create anchors, how the students used anchors to perform search tasks, and the roles of thinking style and spatial ability on learning processes and performance. The 671 generated anchors were categorized as direct, indirect, symbolic, temporal, spatial, and challenging. According to our results, students with legislative thinking style tendencies created the largest number of anchors but rarely used them for subsequent search tasks; executive-style students tended to make symbolic and temporal spatial geographic references and regularly used the teacher-created anchors and the anchors created by judicial-style students were evenly distributed across all categories. In addition, low-spatial ability students tended to create direct geographic references and to use predefined anchors for all problem types.

### Keywords

Anchoring behavior, Thinking style, Spatial ability, Geographic information system, Google Earth

### Introduction

Information technology (IT) is at the center of educational reform and innovation (Dexter, Anderson, & Becker, 1999; Dias, 1999; Mehlinger, 1996; Moursund, 1992), with many countries establishing national guidelines for the incorporation of IT tools into their education systems (Kozma, 2003). Various open source and freeware information products based on the Web 2.0 co-creating and sharing model provide teachers with a range of tools for curriculum design and instruction. Google Earth (<http://earth.google.com>), a free three-dimensional geographic information system (GIS) released in June of 2005, offers innovative approaches to studying geography and earth science topics using simulated images (Doering & Veletsianos, 2007; Lund & Macklin, 2007; Patterson, 2007; Rutherford, 2008; Sherman-Morris, Morris, & Thompson, 2009). An example of a Google Earth user interface is shown in Figure 1.

In addition to providing support for situated learning, Google Earth can also be used as a platform for knowledge integration among multiple domains. Users can create place marks, add text and graphics to those place marks, save them, designate them as favorites, and share them. These functions can serve as anchors for student-written notes containing geographic references while students explore domains consisting of geography-based information. Students can apply anchors and place marks to multiple cognitive functions in order to cultivate creative thinking. Here we will describe our proposal for a novel instructional approach that employs Google Earth for learning concepts in geography, earth science, and history.

We believe that multidisciplinary learning activities associated with Google Earth can help students construct and integrate knowledge from multiple domains, but further research is required to determine the positive uses of GISs in learning to make sure that students do not get lost in multidimensional simulated spaces and that knowledge integration across multiple domains does not hinder learning. To address these issues, we observed student anchoring activities such as referencing and citing, and examined the effects of different thinking styles and spatial abilities on anchoring activities.

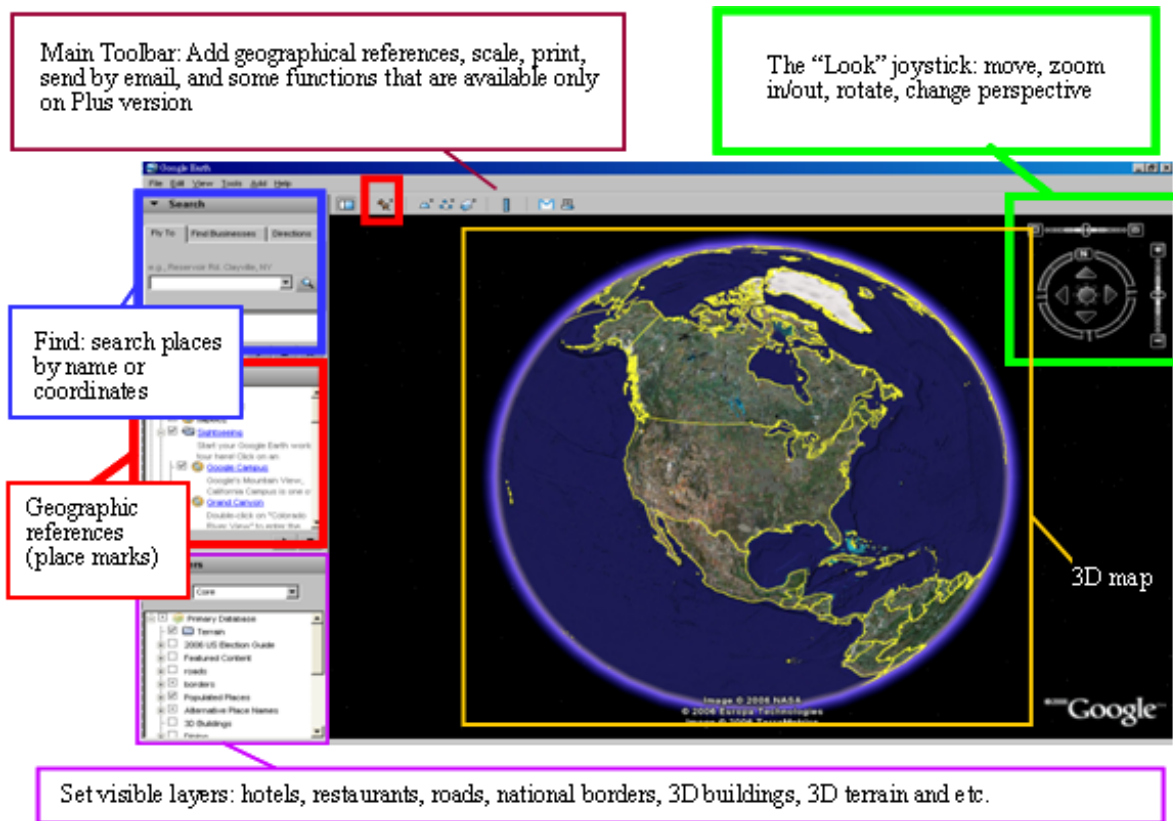


Figure 1. Screenshot of Google Earth user interface

## GIS anchoring

Kienreich, Granitzer, and Lux (2006) combined a web-based map with a Brockhaus electronic encyclopedia to help readers access geographic references associated with place names, geological principles, historical events, and specific encyclopedic and geographic domains, and named their process geospatial anchoring. We used their work as the basis for our experiment, in which we introduced two exploration activities to a group of Taiwanese high school freshmen: creating anchors and using them to perform search tasks. Google Earth was used to establish a virtual learning context and to support student navigation for exploring world geography while they read web-based articles on earth science and history topics. Students were shown how to create geographic reference placemarks and how to write summaries of article content in placemark description fields. We use the term “anchor” to indicate a placemark that includes a geographic reference and explanatory notes, and “anchoring” to describe the act of creating and adding ideas to Google Earth placemarks. An example of a Google Earth anchor is shown in Figure 2. In the following activity, some anchors were predefined by the teacher for subsequent search applications. Students were given the option of using anchors to perform geographic searches or to otherwise use their geographic knowledge to find target images by rotating the Google Earth globe. We used our observations of these activities to investigate how anchoring behaviors can support knowledge integration.

The five types of geographic references defined by Kienreich, Granitzer, & Lux (2006) are direct, indirect, symbolic, temporal spatial, and challenging. Direct geographic references link a location with a place specifically mentioned in a text. One example of an indirect geographic reference is the association between the Forbidden City Museum and Beijing. Symbolic references connect geographic objects with a set of objects mentioned in an article—for example, a Greenland anchor may be linked to an article on glaciers. Temporal spatial references are tied to historical events such as wars and scientific discoveries. Challenging references involve links requiring additional information. In the present study, direct and indirect references required the application of geographic knowledge, symbolic references focused on earth science knowledge (e.g., geology, climate), temporal spatial references required knowledge of

historical events, and responses to challenging geographic references required multiple sources of knowledge across several domains.

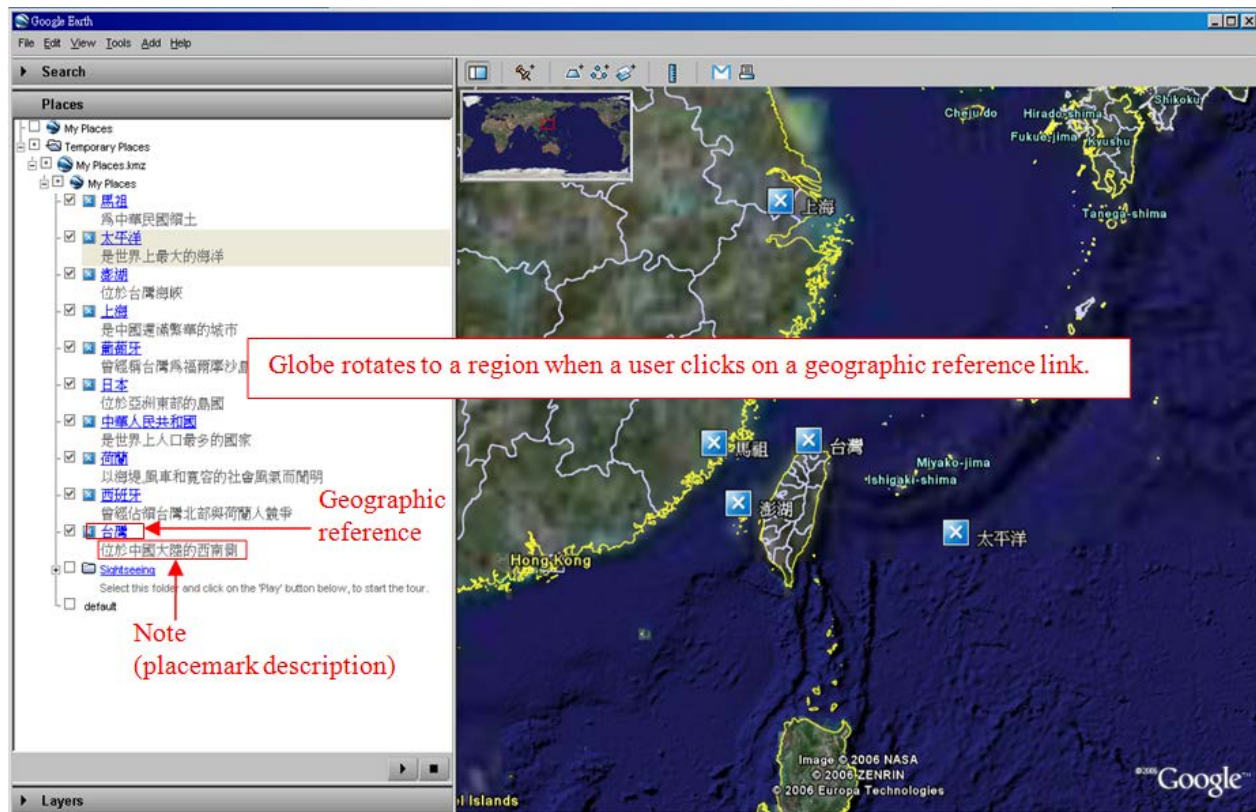


Figure 2. Google Earth anchor example

### Spatial ability and learning functions

According to Greenfield (1984), navigation performance in a virtual space is largely dependent on spatial ability. This competency is generally believed to have two components: orientation and visualization (McGee, 1979). Lohman (1979) defines spatial ability as the ability to generate, retain, retrieve, and transform well-structured visual images, while Pellegrino and Hunt (1991) define it as the ability to reason about movement in a visual field. Gorgorió (1998) has proposed the concept of spatial processing ability, which she defines as the ability to fulfill the mix of mental operations required to solve spatial tasks. This includes skills tied to imagining and visually decoding spatial objects, relationships, and transformations, plus the ability to encode them verbally. Schofield and Kirby (1994) define spatial ability as the ability to visualize spatial relationships, including transforming objects from 3D to 2D, and identifying shapes in two-dimensional spaces. Google Earth users view a three-dimensional globe from the perspective of an orbiting astronaut. The ability to move, identify, locate, compare, rotate, and transform objects in 3D entails spatial relations, visualization, and orientation processes. Our assumption is that learners' behaviors in 3D GISs are tied to their ability to create and work with mental images—a type of spatial ability.

Thinking style is an important aspect of the learning process (Zhang & Sternberg, 2005). The list of researchers who have illustrated the significant contributions of thinking style to academic achievement in traditional environments includes Bernardo, Zhang, and Callueng (2002), Cano-Garcia and Hughes (2000), Grigorenko and Sternberg (1997), Zhang (2008), and Zhang and Sternberg (2005). Fan, Zhang, and Watkins (2010) have done the same for hypermedia learning environments. Thinking style represents personal tendencies and attitudes toward using one's abilities (Sternberg, 1994, 1997). According to Sternberg, it is closely associated with mental processes and preferences, with all individuals having a preferred thinking style for dealing with life problems. He lists the five dimensions of mental self-government as function, form, level, scope, and leaning, and the three function categories as legislative (preferring one's own way of doing things), executive (concerned with accomplishing assigned tasks within a set of

guidelines), and judicial (tending to evaluate rules and processes, and preferring problems that require analysis). During learning activities, students generally set goals according to their individual habits and thinking styles. Therefore, we predicted that thinking style would influence our participants' anchoring activities.

## **Study purpose**

For this investigation we defined the five above-described anchor types and considered the effects of learner thinking style and spatial ability on anchoring behaviors and learning performance. In addition to the anchoring activity, we designed a search activity to determine whether students used anchors as support tools or shortcuts for task completion, and whether learning outcomes from the first activity affected subsequent search tasks. Our goal was to determine if and how anchoring behaviors during learning activities are affected by individual differences in spatial ability and thinking style. In this study, we used three thinking styles (legislative, executive, and judicial) and three levels of spatial ability (high, medium, and low) as independent variables and two types of anchoring behaviors (creating and using anchors) as dependent variables. We established five research questions:

- Is anchor creation in a GIS activity affected by thinking style?
- Is anchor creation in a GIS activity affected by level of spatial ability?
- Is the use of anchors in a GIS activity affected by thinking style?
- Is the use of anchors in a GIS activity affected by level of spatial ability?
- Is search performance in a GIS activity affected by thinking style, spatial ability, or anchoring behaviors?

## **Method**

### **Participants and measures**

The final sample of 66 Taiwanese high-school freshmen was established from an original sample of 237 students. Selections were made based on results from Sun and Wang's (2004) Thinking Styles Questionnaire for Junior High School Students, a modification of Sternberg's (1997) Thinking Styles Inventory. The 24 items were designed to identify legislative, executive, and judicial thinking styles (eight each). An example of legislative style is "I like to deal with academic problems that I can use my own methods to solve"; an example of executive style is "I prefer work tasks or problem-solving according to clear rules and operating guidelines"; an example of judicial style is "I like to analyze, evaluate, and compare different views." All responses were given along a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Internal consistency coefficients ( $\alpha$ ) for the questionnaire and its three dimensions have been measured as 0.83 overall, 0.75 legislative, 0.73 executive, and 0.74 judicial (Sun & Wang, 2004). All students in the final sample had high scores in one style and low to medium scores in the other two.

To examine the effects of different spatial processing abilities on anchoring behavior, we ranked the 66 students according to their scores on a spatial ability test developed by Lu, Lu, and Ou (1994) (0.76 internal consistency coefficient). The final groups consisted of 20 low, 26 mid, and 20 high-spatial ability students. In the spatial ability tests, students were asked to choose one of four graphs that most resembled the original; the direction of the original graph was different from the directions of the other four.

The final sample consisted of 40 male students and 26 female students. The majority of participants stated that they had approximately seven years of experience using computers, mostly for Internet surfing, email, and instant chat communication. None had any experience with a learning activity based on Google Earth.

### **Procedure**

The experiment was conducted over a four-week period. During Week 1, thinking style and spatial ability tests were administered during a single 45-minute class period. To increase system familiarity and to reduce the impact of insufficient computer experience, during Week 2 we gave participants 45 minutes of anchoring training and Google Earth instruction. The first activity was assigned during a 45-minute period in Week 3. Students were asked to view maps and create anchors according to article content. English translations and geographic details for major locations were given in the assigned articles to reduce interference due to language ability or background knowledge. The



second activity was done during a 45-minute period in Week 4; during this session, students were asked to complete fifteen search tasks. Here we will describe the two activities in detail:

- *Locate geographic references based on Wikipedia article excerpts.* An example of geographic content is “Taiwan is located in eastern Asia, western Pacific”; an earth science example is “Taiwan is an island extruded and uplifted by the Eurasian and Philippine Sea plates”; a history example is “From 1626 to 1642, northern Taiwan was occupied by Spain.” We purposefully provided articles about familiar locations and basic geographic and historical knowledge in an effort to reduce the effects of individual differences in prior knowledge. Students were instructed to read the articles, use the content to create placemarks (map locations), and add brief explanations to each placemark. They were permitted to rotate the Google Earth globe to perform searches, and to use English names or coordinates to access locations of interest. The purpose of this task was to determine how the students connected textual content to geographic location and how they created geographic references on a map. Anchor numbers and types were evaluated to determine which kinds of subject knowledge were easily discovered and recorded. Anchors were classified according to content; an example of a symbolic geographic reference/anchor would be a placemark on Jade Mountain with the descriptor “highest mountain in Taiwan.” Off-topic anchors or anchors with incorrect explanations were considered invalid and excluded. Students were informed that their second search activity would be based on the same article as in the first; therefore, they could customize their anchors with their own notes.
- *Perform multiple photograph search tasks.* During their searches, students could choose between using one of 30 predefined geographic references/anchors or referring to details provided in the search problem texts. Solutions based on other software functions or methods were disallowed. To determine which anchor types were used most often, we selected five predefined anchors that corresponded to the five problem types: direct, indirect, symbolic, temporal-spatial, and challenging (three of each type, for a total of 15 problems). In all, 30 predefined anchors were established by an instructor prior to the experiment. While predefined anchors may have been historically or geographically associated with correct answers, they did not necessarily contain correct answers. Anchors mostly served as shortcuts for students to quickly move to target areas. Students were also asked to upload map screenshots that matched targets in terms of direction, distance, and angle.

## Scoring

As part of activity 1, students uploaded the anchors they created as (keyhole markup language zipped) kmz files. Figure 2 presents a list of all anchors created by one student, categorized as direct, indirect, symbolic, temporal spatial, or challenging based on Kienreich et al.’s (2006) geographic reference definitions; the same results are presented in English as Table 1. Activity 2 scores reflect differences between target photos and search results. The highest possible score of 5 indicated that a student’s search results completely matched the target in terms of direction, distance, and angle. Scores for all problems were summed; a scoring example is shown in Figure 3.

*Table 1. Anchor list for student S001*

No.	Anchor Name	Anchor Note	Anchor Type
1	Matsu	Republic of China territory	Direct
2	Pacific Ocean	World’s largest ocean	Symbolic
3	Penghu	In the Taiwan Strait	Direct
4	Shanghai	A bustling city in China	Indirect
5	Portugal	Home of Portuguese, who called Taiwan “Formosa”	Temporal spatial
6	Japan	An island country in East Asia	Symbolic
7	People’s Republic of China	Most populous country in the world	Challenging
8	Netherlands	Known for dikes, windmills, and a tolerant society	Challenging
9	Spain	Once occupied northern Taiwan and competed with the Dutch	Temporal spatial
10	Taiwan	Located southwest of China	Indirect

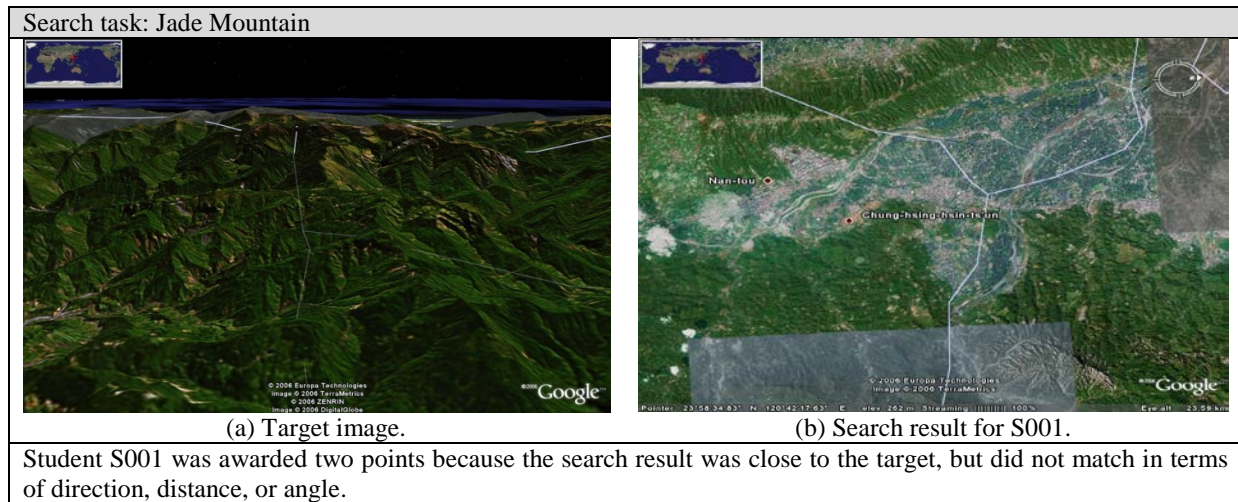


Figure 3. Image search scoring example

## Results

### Effects of thinking style on creating anchors

During the first activity, participants created 111 direct, 86 indirect, 161 symbolic, 174 temporal spatial, and 85 challenging anchors; anchor numbers according to thinking style are presented in Table 2. Chi-square test results indicate a significant relationship between anchoring behavior and thinking style ( $\chi^2(8) = 44.520$ ,  $p < .001$ ). According to the data on relationships between thinking style and anchoring behavior, legislative-style students created the largest number of anchors and showed preferences for temporal spatial, direct, and symbolic anchors, likely reflecting their enthusiasm for discovering new knowledge (Fig. 4a). Executive-style students created greater numbers of symbolic and temporal spatial anchors. Statistical significance was not found for the main effect of judicial thinking style. Among judicial-style students, anchors were evenly distributed across all categories.

Table 2. Anchor type according to thinking style ( $n = 66$ )

Thinking style	Anchor Type					Overall	$\chi^2(\text{column})$ d = 4
	Direct	Indirect	Symbolic	Temporal spatial	Challenging		
Legislative (n = 22)	61	39	55	68	37	260	14.231**
Executive (n = 22)	18	16	70	63	14	181	85.436***
Judicial (n = 22)	32	31	36	43	34	176	2.580
Overall	111	86	161	174	85	617	56.736***
$\chi^2(\text{row})$ d = 2	26.000***	9.512**	10.820**	6.034*	11.035**	21.592***	

\* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$

### Effect of spatial ability on anchoring

Results for numbers of anchors created by the participants according to spatial ability are presented in Table 3. A significant relationship was found between spatial ability and anchoring behavior ( $\chi^2(8) = 65.175$ ,  $p < .001$ ), and a strong main effect of spatial ability on anchoring was also observed ( $\chi^2(2) = 17.906$ ,  $p < .001$ ). Overall, students in the mid-spatial ability group created larger numbers of anchors compared to the high- and low-spatial ability students



(Fig. 4b). Our results also indicate that students in the low-spatial ability group preferred direct anchors, mid-spatial ability students favored symbolic and temporal spatial anchors, and high-spatial ability students preferred symbolic anchors. No significant differences were noted among these groups in terms of indirect anchor numbers.

Table 3. Anchor type according to spatial ability ( $n = 66$ )

Spatial ability	Anchor Type					Overall	$\chi^2(\text{column})$ d = 4
	Direct	Indirect	Symbolic	Temporal spatial	Challenging		
Low (n = 20)	63	24	28	42	20	177	34.667***
Mid (n = 26)	35	30	68	83	39	255	42.235***
High (n = 20)	13	32	65	49	26	185	44.595***
Overall	111	86	161	174	85	617	56.736***
$\chi^2(\text{row})$ d = 2	33.946***	1.209	18.497***	16.586***	6.659*	17.906***	

\* $p < .05$ , \*\*\* $p < .001$

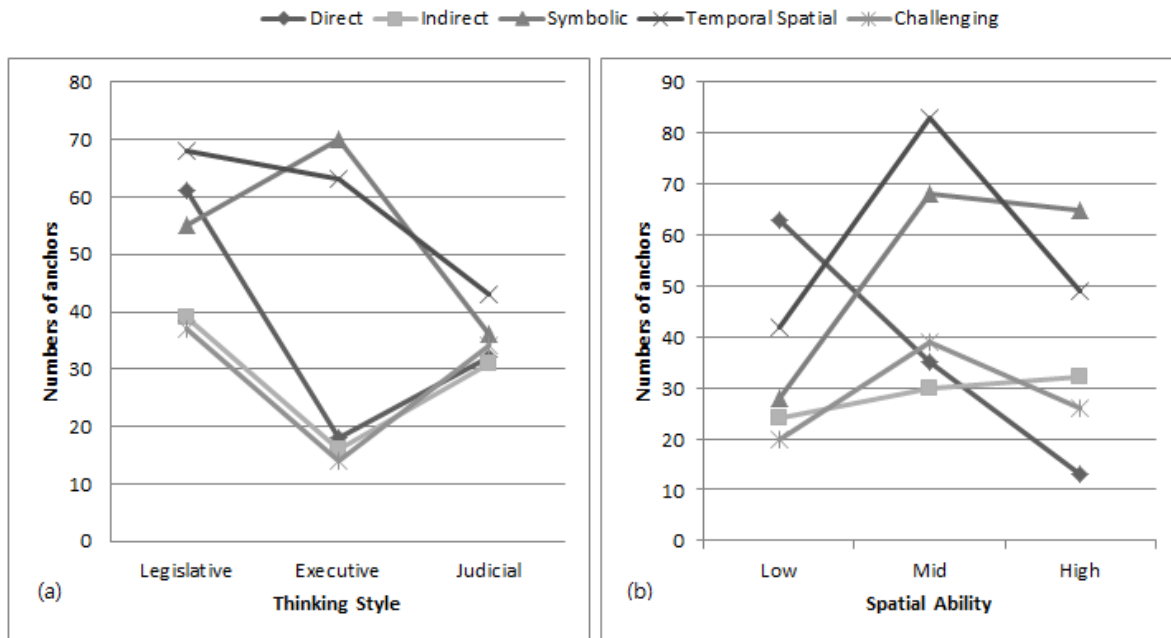


Figure 4. (a) Relationships among thinking styles and numbers of created anchors; (b) Relationships among spatial ability levels and numbers of created anchors

### Effects of thinking style on anchor usage

During the search activity, the 66 participants used predefined anchors 78 times for direct geographic reference problems, 95 times for indirect problems, 130 times for symbolic problems, 101 times for temporal spatial problems, and 143 times for challenging problems. We performed a one-way ANOVA to analyze the average number of problems (out of 15) that were solved with the help of predefined anchors. Our results show that executive-style students ( $M = 11.05$ ) used more anchors than judicial ( $M = 7.45$ ) or legislative-style students ( $M = 6.36$ ). In other words, executive-style students (who tend to conform to rules and prescribed processes) showed the strongest preference for predefined anchors. Results from Chi-square tests examining the use of predefined anchors are shown in Table 4. No statistically significant relationships were found between anchor usage and any of the three thinking

styles ( $\chi^2(8) = 4.145, p = .844$ ), but significance was noted for the main effects of thinking style ( $\chi^2(2) = 31.857, p < .001$ ) and problem type ( $\chi^2(4) = 25.751, p < .001$ ). These data also indicate that executive-style students showed a preference for predefined anchors, but they also indicate that all students showed the same preference when addressing challenging and symbolic problems. Specifically, the results indicate that executive-style students tended to use predefined anchors for all problem types, legislative-style students were more likely to use them to solve symbolic and challenging problems, and judicial-style students were more likely to use them to solve challenging geographic reference problems.

Table 4. Predefined anchor usage according to thinking style ( $n = 66$ )

Thinking style	Problem Type					Overall	$\chi^2(\text{column})$ d = 4
	Direct	Indirect	Symbolic	Temporal spatial	Challenging		
Legislative (n = 22)	18	23	38	23	38	140	12.500*
Executive (n = 22)	39	43	53	50	58	243	4.798
Judicial (n = 22)	21	29	39	28	47	164	12.707*
Overall	78	95	130	101	143	547	25.751***
$\chi^2(\text{row})$ d = 2	9.923**	6.653*	3.246	12.257**	4.210	31.857***	

\* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$

### Effects of spatial ability on anchor usage

No statistically significant relationships were noted between spatial ability and the use of anchors for any problem type ( $\chi^2(8) = 1.656, p = .990$ ). However, the results shown in Table 5 indicate statistical significance for the main effect of spatial ability ( $\chi^2(2) = 23.74, p < .001$ ), suggesting that low-spatial ability students had a stronger preference for predefined anchors compared to students in the other two groups. No significance was noted for the simple main effect of low-spatial ability, meaning that students in this category did not show a preference for any problem type and tended to use predefined anchors for all problem types. Statistical significance was also noted for the simple main effects of mid and high-spatial ability ( $\chi^2(4) = 10.131, p < .05, \chi^2(4) = 11.000, p < .05$ ). Students in these two groups preferred using predefined anchors to solve challenging and symbolic geographic reference problems. For direct and symbolic geographic reference problems, mid- and low-spatial ability students showed strong preferences for predefined anchors. None of the groups expressed significant preferences for using anchors to solve indirect, temporal spatial, or challenging geographic reference problems. Students in the high-spatial ability group tended to use anchors to solve challenging and symbolic geographic reference problems.

Table 5. Predefined anchor usage according to spatial ability ( $n = 66$ )

Spatial Ability	Problem Type					Overall	$\chi^2(\text{column})$ d = 4
	Direct	Indirect	Symbolic	Temporal Spatial	Challenging		
Low (n = 20)	34	39	52	41	53	219	6.365
Mid (n = 26)	28	35	48	35	52	198	10.131*
High (n = 20)	16	21	30	25	38	130	11.000*
Overall	78	95	130	101	143	547	25.751***
$\chi^2(\text{row})$ d = 2	6.462*	5.642	6.338*	3.881	2.951	23.740***	

\* $p < .05$ , \*\*\* $p < .001$

## Effects of thinking style, spatial ability, anchoring behavior, and use of anchors on search performance

The average search task score for all of the participating students was 57.85 (maximum 75) ( $SD = 8.663$ ; skewness =  $-.573$ ). Results from a one-way independent sample ANOVA indicate that executive ( $M = 61.68$ ) and judicial style ( $M = 59.32$ ) students performed significantly better than legislative-style students ( $M = 52.55$ ) ( $F = 8.015$ ,  $p < .01$ ). According to these data, students who tended to strictly follow rules or to carefully choose the proper time to use an anchor performed better overall, but no statistically significant differences were found among students at different spatial ability levels.

We conducted Pearson correlation analyses to examine whether student search performance in the second activity was affected by anchoring behavior in the first activity or the use of predefined anchors in the second activity. As shown in Table 6, search performance was not significantly associated with the number of anchors created in activity 1, with the exception of students who preferred creating direct anchors. However, positive relationships were found between student performance on search tasks and the frequency of using predefined indirect, temporal spatial, and challenging anchors, suggesting that proper guidance and navigation enhanced learning effectiveness.

Table 6. Results from Pearson correlations involving anchors, anchor usage, and search performance ( $n = 66$ )

		Anchor Type					Search Task Score
		Direct	Indirect	Symbolic	Temporal Spatial	Challenging	
Problem Type	Direct	.044	-.204	-.129	.060	-.167	.219
	Indirect	.141	-.341**	-.039	-.207	-.163	.300*
	Symbolic	.212	-.256*	-.190	-.102	-.179	.118
	Temporal spatial	-.023	-.206	-.121	.059	-.217	.327**
	Challenging	.040	-.107	-.194	-.226	-.324**	.271*
Number of predefined anchors used		.101	-.279*	-.169	-.104	-.265*	.312*
Search task score		-.399**	-.033	.204	.051	-.041	—

\* $p < .05$ , \*\* $p < .01$

## Discussion

Our goal was to create a method for encouraging students to search for geographic locations based on textual information, an approach that fits well with the constructivist emphasis on establishing learner knowledge through orientation, elicitation, restructuring, application, and review (Driver & Oldham, 1986). Instruction consisted of an anchoring activity and a search activity based on Google Earth, with students exploring places associated with concepts or events—for example, the location of the National Palace Museum and the Spanish occupation of northern Taiwan. The anchoring activity encouraged the extraction of essential concepts from Wikipedia articles, with students restructuring knowledge in ways that reflected their individual learning styles. Students reviewed their anchors and applied them to the search activity. We predicted that students would internalize knowledge during the anchoring and search processes.

We developed open-ended tasks and allowed students to select locations of interest on 3D maps for the purpose of creating geographic references. These references encouraged students to apply and integrate domain knowledge through the use of Google Earth maps. In addition, by placing anchors on 3D maps, students had the opportunity to develop a deeper understanding of article content, thereby completing search tasks more easily. According to Beane's (1997) integrated curriculum concept, this type of situated learning activity benefits students in terms of conceptual understanding and integration. Our results indicate that the students constructed similar numbers of each type of anchor to complete their tasks. For their direct or indirect anchors they emphasized geographic knowledge, for their symbolic anchors they emphasized earth science knowledge, and for their temporal spatial anchors they emphasized history knowledge. This suggests that Google Earth can be used as a platform for knowledge integration, and that the anchoring approach is suitable for teaching geography, earth science, and history concepts.

Taking thinking style into account, our results indicate that executive-style students were the most likely to create symbolic and temporal spatial anchors. Their careful reading of the content of earth science and history articles is characteristic of an executive-learning style, as was their use of knowledge-based notes associated with article content. However, they often cut-and-pasted article content to use as anchor text—a strategy considered a surface approach (Dunkin & Biddle, 1974). Further, they were rarely willing to search for place locations or to create direct or indirect anchors.

Legislative-style students tended to create direct and temporal spatial anchors and to establish much larger numbers of anchors compared to students in the other groups—an indicator of their enthusiasm for exploring new knowledge. Judicial-style students did not show a preference for any single anchor type, perhaps an indicator of their analytical tendencies. Legislative-style and judicial-style students were willing to create challenging anchors that required additional information from knowledge sources across multiple domains. In other words, the legislative- and judicial-style study participants used deep approaches to learning, while executive students tended to use a surface approach—a finding that agrees with results reported by Zhang (2000) and Zhang and Sternberg (2000). Since the task we used in our experiment involved greater cognitive complexity and encouraged students to create their own knowledge, we believe it facilitated the use of a legislative approach to learning.

Our results also indicate a relationship between anchoring behavior and spatial ability. Students with low-spatial ability scores showed a preference for direct anchors—perhaps as a timesaving tactic, or an indication that they found it difficult to deal with spatial relationships. Both mid- and high-spatial ability students favored symbolic and spatial temporal anchors. The higher level of detail in the articles used for this study is a likely explanation for these preferences. Mid-spatial ability students also showed a tendency to create more anchors. A possible explanation for why low-spatial ability students created fewer anchors is their relative lack of skill in fixing positions on maps, and a possible explanation for why high-spatial ability students created fewer anchors is their perception that recording direct geographical references was unnecessary, since their spatial abilities were sufficient for meeting the requirements of the location search test.

Regarding anchor-citing behaviors, we found that executive-style students were more likely than others to use predefined anchors for search tasks, perhaps due to their characteristics of following rules and finding practical solutions to problems (Sternberg, 1997). Among more analytical (judicial-style) students, their use of anchors was consistent across all problem types. We also noted that legislative-style students were more creative and attentive when using Google Earth, and were therefore less likely to use predefined anchors. Low-spatial ability students clearly had difficulty rotating the Google Earth globe when searching for targets and therefore relied more on predefined anchors. As expected, high-spatial ability students enjoyed rotating the globe to perform their searches, suggesting that they were more likely to benefit from such a system for learning earth science and geography concepts. In short, the data suggest that the executive-style and low-spatial ability students required more structured learning aids and scaffolding when using the 3D GIS.

In terms of search performance, students who showed preferences for predefined anchors performed better on search tasks, suggesting that anchors significantly enhanced search performance to the degree that it was not affected by spatial ability. A possible explanation is the fixed nature of search task solutions versus potential variation in strategy. Accordingly, existing and available anchors may have facilitated searches in the same manner as shortcuts. A standard approach to answer assessment may be supportive of an executive style; our results indicate that executive-style students outperformed their legislative-style classmates. This finding is in agreement with reports from Cano-Garcia and Hughes (2000) and Zhang and Sternberg (1998) of positive correlations between executive thinking style and achievement. In Cano-Garcia and Hughes's study, students who explicitly stated their aversion to creating and planning (negative legislative) and those who favored existing rules and procedures (executive) achieved higher scores. In the present study, even though judicial-style participants used fewer predefined anchors, they earned higher scores on their search performance tasks. This unexpected finding, which suggests that judicial-style students were more analytical, careful, and precise in their use of available anchors, is an example of Zhang's (2004) observation of mixed ways in which judicial-style tendencies contribute to academic achievement.

According to Sternberg and Zhang (2005), "What is valued in one time and place may not be valued in another". According to our data, executive-style students were very capable of organizing article knowledge for use with Google Earth, but less capable of constructing original knowledge. Still, they were skillful at finding landscape photographs via various anchor types that served as task completion shortcuts. In contrast, legislative style students

were more active in creating multidisciplinary knowledge anchors, suggesting that classroom instruction might benefit from teacher efforts to match executive- and legislative-style students for activities involving collaborative knowledge construction. In accordance with social constructivist principles that encourage learner discussion, negotiation, and coordination for integrating new information (Resnick, 1989), we believe that matching these types of students for lessons involving Google Earth-like activities may lead to enhanced knowledge construction.

## Conclusions

GIS activities have been used in teaching curricula for subjects such as archaeology (Hespanha, Goodchild, & Janelle, 2009), sociology (Grady, 2007), economics (Booker, 2007), and criminology (Harries, 1999). For the present study, we used a free GIS application, Google Earth, to create a teaching scenario in which learners were asked to organize article content in terms of geographical references and to present their results on 3D maps. The goal of the activity was to help students establish connections between geographic locations and article content, thus building a deeper understanding of the meaning and value of specific knowledge. In such scenarios, instructors serve as navigators and facilitators, while students act as explorers and knowledge producers. Google Earth supports learning in a manner that matches Dias and Atkinson's (2001) claim regarding characteristics for integrating technology into curricula, as well as Moersch's (1999) fourth-level integration of technology into classrooms. We believe this activity can be used to help establish critical thinking and knowledge construction skills leading to higher-order learning. According to our study results, learner use of Google Earth's anchoring feature is affected by a combination of thinking style and spatial ability. Further, we found little variation in terms of spatial ability, which can be enhanced through cumulative experience in the use of 3D GISs or maps. In summary, we found that thinking style exerted a stronger influence than spatial ability on learning among the study participants.

Teachers play key roles in the integration of IT tools into learning environments (Hsu & Sharma, 2008; Jung, 2005; Vanderlinde & van Braak, 2011). They are expected to facilitate learning and to make learning meaningful for individual students rather than to simply provide knowledge and skills (Roblyer, 2006). Students with different thinking styles tend to prefer specific teaching approaches (Zhang, 2004). For example, legislative- and judicial-style students prefer approaches that emphasize conceptual change, and executive-style students prefer approaches that emphasize information transmission. In their roles as facilitators and advisors, teachers can benefit from emphasizing learning processes rather than learning outcomes, and from designing materials and learning contexts that match the needs of a range of learning styles. They can support their students' self-learning efforts by providing them with greater decision-making autonomy. Tools such as Google Earth can be used to observe student anchoring behaviors and to identify how they use anchors to construct knowledge networks; teachers can then use this information to incorporate anchoring activities into learning, thereby transforming unidirectional and static learning processes into those in which pluralistic, life-oriented, and interesting experiences are accumulated (Jonassen, 1996; Sprague & Dede, 1999).

## Limitations

Anchoring in Google Earth requires strong IT skills and experience. Therefore, the instructional activities described in this paper may not be suitable for students who are not proficient in the use of computers. Another potential limitation concerns the article topics, since familiarity with content likely affected the anchoring behaviors of the study participants. The articles used in this study were on geography, earth science, and history topics, which may have influenced student decisions to create geographic, symbolic, and temporal spatial anchors. Students may express different anchoring preferences for other disciplines. Finally, we asked a teacher to establish 30 anchors in advance to support our observations of predefined anchor usage. The specific ways that students respond to anchors created by others requires further exploration.

## Acknowledgements

The authors would like to thank the National Science Council of Republic of China for financially supporting this research under contract numbers NSC100-2511-S-126-006-MY2 and 99-2511-S-009-009-MY3.

## References

- Beane, J. A. (1997). *Curriculum integration-designing the core of democratic education*. New York, NY: Teachers College, Columbia University.
- Bernardo, A. B., Zhang, L. F., & Callueng, C. M. (2002). Thinking styles and academic achievement among Filipino students. *The Journal of Genetic Psychology*, 163, 149–163.
- Booker, J. (2007) Economics: Exploring spatial patterns from the global to the local. In D. S. Sinton & J. J. Lund (Eds.), *Understanding place: GIS and mapping across the curriculum* (pp. 97–110). Redlands, CA: ESRI Press.
- Cano-Garcia, F., & Hughes, E. H. (2000). Learning and thinking styles: An analysis of their relationship and influence on academic achievement. *Educational Psychology*, 20(4), 413–430.
- Dexter, S. L., Anderson, R. E., & Becker, H. J. (1999). Teachers' views of computers as catalysts for changes in their teaching practice. *Journal of Research on Computing in Education*, 31(3), 221–239.
- Dias, L. B. (1999). Integrating technology: Some things you should know. *Learning and Leading with Technology*, 27(3), 10–13.
- Dias, L. B., & Atkinson, S. (2001). Technology integration: Best practices—Where do teachers stand? *International Electronic Journal for Leadership in Learning*, 5(11). Retrieved from <http://iejll.synergiesprairies.ca/iejll/index.php/ijll/article/view/503>
- Doering, A., & Veletsianos, G. (2007). Authentic learning with geospatial data: An investigation of the use of real-time authentic data with geospatial technologies in the K–12 classroom. In R. Carlsen et al. (Eds.), *Proceedings of society for information technology & teacher education international conference 2007* (pp. 2187–2193). Chesapeake, VA: AACE.
- Driver, R., & Oldham, V. (1986). A constructivist approach to curriculum development in science education. *Studies in Science Education*, 13, 105–22.
- Dunkin, M. J., & Biddle, B. J. (1974). *The study of teaching*. New York, NY: Holt, Rinehart & Winston.
- Fan, W., Zhang, L. F., & Watkins, D. (2010). Incremental validity of thinking styles in predicting academic achievements: An experimental study in hypermedia learning environments. *Educational Psychology*, 30(5), 605–623.
- Gorgorió, N. (1998). Exploring the functionality of visual and non-visual strategies in solving rotation problems. *Educational Studies in Mathematics*, 35, 207–231.
- Grady, J. (2007) Sociology: Surprise and discovery exploring social diversity. In D. S. Sinton & J. J. Lund (Eds.), *Understanding place: GIS and mapping across the curriculum* (pp. 83–95). Redlands, CA: ESRI Press.
- Greenfield, P. M. (1984). *Mind and Media: The Effects of Television, Video Games, and Computers*. London, UK: Fontana Paperbacks.
- Grigorenko, E. L., & Sternberg, R. J. (1997). Styles of thinking, abilities, and academic performance. *Exceptional Children*, 63(3), 295–312.
- Harries, K. (1999) *Mapping Crime: Principle and Practice*. Washington, DC: United States Department of Justice.
- Hespanha, S. R., Goodchild, F., & Janelle, D. G. (2009). Spatial thinking and technologies in the undergraduate social science classroom. *Journal of Geography in Higher Education*, 33(S1), S17–S27.
- Hsu, P. S., & Sharma, P. (2008). A case study of enabling factors in the technology integration change process. *Educational Technology & Society*, 11(4), 213–228.
- Jonassen, D. H. (1996). *Computers in the Classroom: Mindtools for Critical Thinking*. Englewood Cliffs, NJ: Prentice Hall.
- Jung, I. (2005). ICT-pedagogy integration in teacher training: Application cases worldwide. *Educational Technology & Society*, 8(2), 94–101.
- Kienreich, W., Granitzer, M., & Lux, M. (2006). Geospatial anchoring of encyclopedia articles. *Proceedings of Tenth International Conference on Information Visualisation (IV'06)*, 211–215. doi:10.1109/IV.2006.57
- Kozma, R. (Ed.) (2003). *Technology, Innovation, and Educational Change: A Global Perspective*. Eugene, OR: International Society for Technology in Education.
- Lohman, D. F. (1979). *Spatial Ability: Individual Differences in Speed and Level*. (Technical Report No. 9. Aptitude Research Project). Stanford, CA: Stanford University.
- Lu, C. Y., Lu, C. M., & Ou T. H. (1994). *Multifactor Aptitude Test*. Taipei, Taiwan: Chinese Behavioral Science Corporation.

- Lund, J. J., & Macklin, S. (2007). ArcGIS and Google Earth: Rules of engagement. *ESRI Education Users Conference*, EDUC 1546. Retrieved February 6, 2010, from [http://proceedings.esri.com/library/userconf/educ07/educ/papers/pap\\_1546.pdf](http://proceedings.esri.com/library/userconf/educ07/educ/papers/pap_1546.pdf)
- McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychological Bulletin*, 85(5), 889–918.
- Mehlinger, H. D. (1996). School reform in the information age. *Phi Delta Kappan*, 77 (6), 400–407.
- Moersch, C. (1999). Assessing current technology use in the classroom: A key to efficient staff development and technology planning. *Learning & Leading with Technology*, 26 (8), 40–43, 49.
- Moursund, D. (1992). *The Technology Coordinator*. Eugene, OR: International Society for Technology in Education.
- Patterson, T. (2007). Google Earth as a (not just) geography education tool. *Journal of Geography*, 106(4), 145–152.
- Pellegrino, J. W., & Hunt, E. B. (1991). Cognitive models for understanding and assessing spatial abilities. In H. Rowe & J. Biggs (Eds.), *Intelligence: Reconceptualization and measurement*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Resnick, L. (1989). *Knowing, Learning, and Instruction*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Roblyer, M. D. (2006). *Integrating Educational Technology into Teaching* (4<sup>th</sup> ed.). Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.
- Rutherford, S. (2008). Earth science education matters! A master's degree program for in-service teachers. *Journal of Geoscience Education*, 56(5), 378–382.
- Schofield, J., & Kirby, J. (1994). Position location on topographical maps: Effects of task factors, training, and strategies. *Cognition and Instruction*, 12(1), 35–60.
- Sherman-Morris, K., Morris, J., & Thompson, K. (2009). Introducing teachers to geospatial technology while helping them to discover vegetation patterns in Owens Valley, California. *Journal of Geoscience Education*, 57(1), 64–72.
- Sprague, D., & Dede, C. (1999). It I teach this way, am I doing my job? Constructivism in the classroom. *Learning & Leading with Technology*, 27(1), 6–9, 16–17.
- Sternberg, R. J. (1994). Allowing for thinking styles. *Educational Leadership*, 52(3), 36–40.
- Sternberg, R. J. (1997). *Thinking Styles*. New York, NY: Cambridge University Press.
- Sternberg, R. J., & Zhang, L. F. (2005). Styles of thinking as a basis of differentiated instruction. *Theory into Practice*, 44(3), 245–253.
- Sun, C. T., & Wang, P. I. (2004). *The learning, affective, and interactive effects of junior high school cooperative groups formed by thinking styles in Learning Powerpoint project*. (Unpublished master's thesis). National Chiao Tung University, Hsinchu.
- Vanderlinde, R., & van Braak, J. (2011). A new ICT curriculum for primary education in Flanders: Defining and predicting teachers' perceptions of innovation attributes. *Educational Technology & Society*, 14(2), 125–135.
- Zhang, L. F., & Sternberg, R. J. (1998). Thinking styles, abilities, and academic achievement among Hong Kong university students. *Educational Research Journal*, 13(1), 41–62.
- Zhang, L. F. (2000). Relationship between thinking styles inventory and study process questionnaire. *Personality and Individual Differences*, 29, 841–856.
- Zhang, L. F. (2004). Do university students' thinking styles matter in their preferred teaching approaches? *Personality and Individual Differences*, 37, 1551–1564.
- Zhang, L. F. (2008). Preferences for teaching styles matter in academic achievement: Scientific and practical implications. *Educational Psychology*, 28(6), 615–625.
- Zhang, L. F., & Sternberg, R. J. (2000). Are learning approaches and thinking styles related? A study in two Chinese populations. *The Journal of Psychology*, 134(5), 469–489.
- Zhang, L. F., & Sternberg, R. J. (2005). A threefold model of intellectual styles. *Educational Psychology Review*, 17(1), 1–53. doi: 10.1007/s10648-005-1635-4.

## Am I Extravert or Introvert? Considering the Personality Effect Toward e-Learning System

Amal Al-Dujaily<sup>1</sup>, Jieun Kim<sup>2</sup> and Hokyoung Ryu<sup>2\*</sup>

<sup>1</sup>Department of Management, Marketing, and Information Systems, Dhofar University, Oman // <sup>2</sup>Graduate School of Technology Innovation and Management, Hanyang University, Korea // amal\_aldujaily@du.edu.om // jkim2@hanyang.ac.kr // hryu@hanyang.ac.kr

\*Corresponding author

(Submitted December 07, 2011; Revised August 17, 2012; Accepted August 29, 2012)

### ABSTRACT

A concern of computer-based learning system design is how to accommodate learners' individual differences during learning activities. Previous research suggests that adaptive e-learning systems can effectively address such individual differences and, consequently, they enable more directed tutoring via computer-assisted instruction. In this paper, we explore this assertion, reflecting on the outcomes of two successive experiments that were performed to empirically demonstrate that learners' personality traits might be significant in understanding differences in learning outcomes from using e-learning systems. One hundred and forty-five participants from Computer Science participated in this study. A two-by-two between-subjects factorial study was designed, where the personality traits derived from Myers-Briggs Type Indicator, and the two computer-based learning systems (adaptive vs. non-adaptive) are independent variables. The results suggested that the personality trait that could be indicative of preferred learning styles; in particular extraverted/introverted personal traits significantly influenced learning activity on adaptive e-learning system. A design guideline might be implicative of how the e-learning system suits the learner's personality trait.

### Keywords

Adaptive e-learning systems, Personality trait, Learning styles, Personality trait, MBTI

### Introduction

The use of computers in education has long been considered a viable delivery mode for learning. As many previous studies suggest, the success of a computer-based learning system can be measured by how long it takes a user to reach their learning objectives. However, it is also recognised that the evaluation of a learning system is naturally contingent upon a learner's own perspective (i.e., user experiences), rather than purely technical measures of system quality. Open to question is thus how to attract learners towards computer-based learning and how to sustain their focus and attention over long periods.

Most early computer-based learning systems were only capable of delivering the same content in, to a great extent, a predefined order to all learners, despite the fact that they might have quite different perspectives and preferences in a given learning domain. As a result, many researchers have criticised such systems as inadequate in raising learners' motivation toward the subject matter (e.g., Clark & Mayer, 2008; 2011; Sampson, Karagiannidis, & Kinshuk, 2002). In contrast, there are other studies (e.g., Bowler et al., 2009; Pulford & Sohal, 2006) that have demonstrated that a better way of dealing with this issue would be to develop different learning materials to cater for individual learners. However, to our knowledge, this has not yet been proved practicable.

As new learning technologies evolve, it becomes more feasible to present more customised learning materials to address the differences of each individual without developing many different versions of learning content (e.g., Curilem et al., 2007). For instance, Intelligent Tutoring Systems (ITS) is a broad term referring to any kind of computer-based learning system incorporating artificial intelligence components. Such systems have previously been implemented in various learning domains, e.g., AutoTutor (Graesser et al., 2005).

These have gained significant attention in the field of computer-based learning, with the belief that such approaches may allow new computer-based education design to increase students' motivation and performance in comparison with other instructional methods (Moundridou & Virvou, 2003). All of the examples mentioned above are designed to effectively adapt to the current levels of knowledge and understanding of the user, and guide them toward their learning goals with the most relevant learning content (Brusilovsky, 2001; De Bra et al., 2002).



Obviously, this type of customisation, demonstrated by many ITS, is a positive development in the field of computer-based education. However, there are still many problematic features in the customisation process, partly because one cannot make computer-based learning fully customised, but mostly because it makes significant demands on the designers of e-learning systems to fully understand how differences between learners would impact their learning styles, and how that knowledge can be incorporated into the design of flexible learning environments. That being said, some early studies on traditional e-learning systems that concern themselves with how different learning styles may be associated with learning design are worthy of note (e.g., Clark & Mayer, 2008). Importantly, Kogan (1971) and Messick (1976) earlier observed that learning content design should be congruent with each learner's preferred learning style. In particular, Kogan (1971) concluded that, based on individual differences in cognitive style, the best learning content design could be identified for each individual learner. Many recent studies have also emphasised the relationship between learning style and the use of e-learning systems (e.g., Clark & Mayer, 2011).

However, none of the studies have provided practical methods for employing learning styles in computer-based learning systems design. The central research question in this article is, therefore, to consider the personality effect as a practical approach to measuring differences in learning styles, thereby gaining a deeper awareness of ways to consider personality traits in designing computer-based learning systems.

## Learning and personality

The attempt to incorporate learners' cognitive styles into computer-based learning is hardly new. For instance, Brusilovsky (2001) identified many cognitive features that are able to drive adaptation for individualised learning, shown in Figure 1, though he did not include personality traits in adaptation.

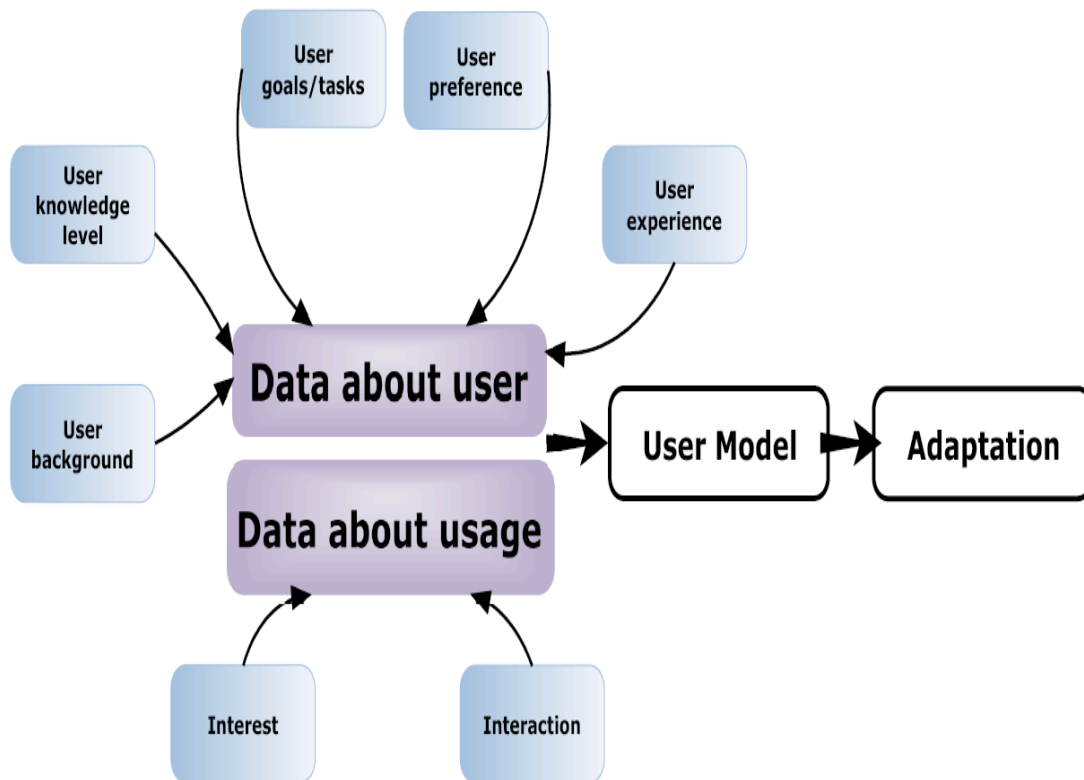


Figure 1. The user model of an adaptive e-learning system, extended from Brusilovsky (2001)

Table 1 also depicts how these cognitive features are considered in intelligent tutoring systems. For instance, ELM-ART (Brusilovsky, Schwarz, & Weber, 1996) provided adaptive navigation support and dynamic content sequencing that were able to match the learner's current goals to their previous knowledge level. The approaches based on the cognitive features described in Table 1 suggest a way forward, but how they can be fully implemented in computer-based learning is still elusive. For instance, the user's current knowledge must be constantly monitored to present the right level of learning content, so many computer-based learning systems have a compulsory test at the end of each learning module. This crude approach is neither universally practical nor optimal in learning experiences (i.e., flow paradox).

*Table 1. User models in adaptive e-learning systems, extended from Brusilovsky (2001)*

	Knowledge level	User goal	Background	Experience	Interaction
Interbook (Eklund & Brusilovsky, 1998)	Continuous monitoring of user's knowledge level	Goals are described by students, which helps them to find a proper learning path.		Prior knowledge of relevant subjects and hypertext experience is collected.	Frames presentation, structure and link annotation are given.
ELM-ART (Brusilovsky et al. 1996a)	Collecting prior example-based problem-solving pattern	Based on a pre-defined learner's goal, students can practise new knowledge by working on exercises.		Adaptive annotation and curriculum sequencing allowed on a user-experience level.	Interactive examples and link annotations are given.
AHA (De Bra & Stash, 2002)	Uses existing adaptation techniques based on learners' knowledge		Collect users' interests, time stamps, and visits to particular web content to adapt to the correct content delivery.		Hide or show webpage links or annotated fragments.
NetCoach/ ART-WEB (Weber et al., 2002)		Lead students through learning materials and monitor learning progress to determine user goals.			Course materials are organised in a tree structure that can be freely browsed by learners.
MLTutor (Smith & Riding, 2003)	Profile-building in order to enable individual adaptation without additional input from the user		Based on applying machine-learning algorithms to generate personalised adaptation to meet user's background	Browsing history to determine user's experience level	Replace user-predefined profiles with a dynamic user profile scheme.

A rather controversial way to cope with individual differences has been proposed by the personality psychologists (e.g., Bayne, 2004; Soles & Moller, 2001). For example, Bayne (2004) claimed that people's personality has a significant influence on how learners may or may not want to become involved in their learning processes, independent of their personal interests or stage of cognitive development. This perspective seems to be of value given the growing complexity of e-learning systems, which contain extensive content and enable various learning paths. Understanding the concept of preferred learning styles to facilitate searching and navigation is a fundamental requirement for adaptive learning systems design, in that such systems need to match a user's learning styles and interests in order to hold their attention in unsupervised learning. Bayne (2004) maintained that personality type

might reflect on the learner's preferences in digesting information and making decisions, bringing the personality effect to the fore in computer-based learning.

There are, however, some other studies that contrast with this deterministic view of personality effect. Mixed outcomes of the personality effect have also been investigated in conjunction with the study of organisational effect, culture, and discipline-related effect, task difficulty, and so on. For instance, Stipek (2001) claimed that many other environmental factors would influence educational practices, so the assumed personality effect might constrain only how a learning motivation is triggered. Allen & Kim (2001) suggested that personality and the context tend to be evolved through their interaction in information-seeking processes.

Though we admit such mixed outcomes of the personality effect, there are two points that make our study worthy of further investigation. First, we still cannot adequately explain why some learners show more enthusiasm and intrinsic desire to know whilst others seem to get easily bored and disinterested. The motivation to learn comes not only from the external content, but also from the self, emphasising the role of the intrinsic nature of the human mind. It is also true that in computer-based learning people must be motivated and rewarded for investing their time and effort to become knowledgeable. In this regard, Khan and Radcliffe (2005) argued that the individual cognitive processing mode (or the traits that come from personality) of each person must match with the information or content presented.

### **Personality and learning styles**

Though there is no consensus of the definition of personality, it is thought that the term refers to dynamic mental structures and coordinated mental processes that determine an individual's emotional and behavioural adjustments to their environment (Millon, 1990). Arguably, according to the theory of conditional reasoning (James & McIntyre, 2000) that provides a conception of learning activity, the reasoning process is highly dependent on personality, that people will frame the same stimuli or events differently. Several general perspectives on learning behaviours thus suggest many grounds for saying that personality relates to an individual's learning process.

First, many researchers (e.g., Deci & Ryan, 1985) have demonstrated that students with a particular trait seem intrinsically motivated, otherwise remaining disengaged. Second, research on individual differences in reasoning styles suggests that those who are high in the trait of achievement motivation consistently direct intense and persistent effort toward accomplishing demanding learning tasks. Finally, research regarding multiple types of intelligence (Gardner, 1983) suggests that students who are taught in a way that matches their reasoning or motivation modes are likely to achieve at higher levels. These views imply that the degree of match between computer-based learning and student personality seems to be an important consideration in adaptive e-learning systems design.

There have been many proposed models, inventories, measures, and questionnaires that help to determine learning styles from personality traits, such as the Big Five theory, Learning Style Inventory (LSI), and the Myers-Briggs Type Indicator (MBTI). Prior research grounded on the Big Five theory all indicates that core aspects of human personality tend to have strong influences on learning behaviours, in particular, motivation and the reasoning process (Costa & McCrae, 1992). Entwistle and Entwistle (1970) found that stable introverts using good study methods achieved higher learning performance than extraverts, whereas Furnham and Medhurst (1995) showed a significant positive correlation between extraversion and performance in a seminar class. Another well-known model of learning style related to personality traits is the Kolb's Learning Style Inventory (LSI). The LSI is comprised of two dimensions in learning process: perception (abstract-concrete) and processing (active-reflective). The combination of these two dimensions dictates that learning style can be classified into convergent learners, divergent learners, assimilators, and accommodators (Kolb, 1984).

Lastly, the Myers-Briggs Type Indicator (MBTI) (Myers et al., 1998) has been widely used and validated extensively in the education domain. MBTI classifies personality traits into four dichotomies: extraverted-introverted, sensing-intuitive, thinking-feeling, and judgement-perception. MBTI has provided many guiding principles to improve learners' performance in learning activity design (e.g., Felder et al., 2002; Soles & Moller, 2001). The evidence for the validity of MBTI theory is, as such, substantial. Many MBTI-based studies from a pedagogical stance are thus worthy of note. For instance, Furnham (1992) found that extraverts tend to be more active, whereas introverts are more reflective. Individuals high in intuitiveness prefer to evaluate information more intuitively, whereas individuals low in this (i.e., sensing people) used systematic information processing. Zhang (2003) demonstrated that thinking and perceiving individuals preferred a deep learning style, whereas feeling and judging ones preferred a superficial

learning style. Also, some studies have examined MBTI personality traits that may be related to aspects of academic motivation. For example, Heaven (1989) reported that achievement motivation was positively correlated with extraversion, and Busato et al., (1999) reported that extraverted students were more achievement-oriented and preferred meaning-, reproduction-, and application-directed learning styles. In contrast, students high on introversion had low achievement motivation, exhibited an undirected learning style, and had difficulty in identifying and processing what material was important.

Interestingly, the learning style theories above are much in common to specify the types of personality. For instance, the two dimensions of Kolb's LSI and MBTI are quite closely replicated: perception (abstract-concrete) in the LSI relates to the feeling-thinking mode of MBTI and processing (active-reflective) is correlated with the extravert-introverted type of MBTI. In a similar manner, four of the five factors in the Big Five theory are on a par with the four dimensions in MBTI, as follows: extraversion (extraversion in the Big Five), sensing (openness to experience in the Big Five), thinking (agreeableness in the Big Five), and judging (correlated to conscientiousness in the Big Five) (McCrae & Costa, 1989). Table 2 presents an overall comparison of the dimensions in the MBTI and the Big Five theory. This convergence between the major theories of personality is encouraging for our study, in that the various approaches, developed in radically different traditions, agree closely on the major characteristics of personality.

*Table 2. Comparing the Myers-Briggs Type Indicator (MBTI) with the Big Five theory*

The Four MBTI Dimensions	Big Five Dimensions					
	Personality Types	Extraversion	Agreeableness	Conscientious	Openness to experience	Emotional stability
	Extraverted-introverted (E-I)	Strongly related to extraversion				Emotional stability is not covered by MBTI
	Sensing-intuitive (S-N)				Related to openness to experience	
	Thinking-feeling (T-F)		Related to agreeableness			
	Judging-perceiving (J-P)			Strongly related to conscientiousness		

### **The current study**

Although prior research has provided a useful lens to see the relationship between various personality traits and selected aspects of academic motivation, few studies have examined the relationship between personality traits and computer-based learning design. The current study was designed to address this by directly assessing the relationship between personality traits derived from MBTI and performance in computer-based learning.

In this paper, we present two studies on the personality effect in computer-based learning. In the following section, we describe the first experiment that was carried out in both Oman and New Zealand, where we examined the impacts of their personality traits as they used different web-based learning systems. Next, the second subsequent experiment was presented and explored a possible answer to the question "why is computer-based learning susceptible to the personality effect?" Based on the findings from both experiments, we empirically drew some implications for designing a computer-based learning system.

### **Experiment I: Personality effect on computer-based learning system use**

The first experimental study was to demonstrate the personality effect (introvert vs. extravert). Hence, only two main manipulations were considered: the personality type (extraversion vs. introversion) and the learning system given

(ITS vs. non-ITS). We sought evidence that a particular learning system induces a different learning performance for different personality types.

## Method

### Participants

A two-way between-subjects experimental design was used. The four experimental groups were formed by factorially combining the two personality types (extraversion vs. introversion) and the two computer-based learning systems (ITS vs. non-ITS), as shown in Table 3. The symbol \* in the table indicates the number of participants from Oman, and the symbol \*\* indicates those from New Zealand. The split in the test groups was due to the fact that the first author of this article worked in both countries to carry out the empirical study. This can be further justified by McCrae's findings (2002) that a person's culture would tell us much less than his or her own personality profile.

*Table 3. The four experimental groups (extraversion vs. introversion/ ITS vs. non-ITS)*

	Extraversion	Introversion	Total
ITS (ELM-ART)	8(4*/4**)	12 (9*/3**)	20 (13*/7**)
Non-ITS (Web)	12(8*/4**)	28 (19*/9**)	40(27*/13**)
Total	20 (12*/8**)	40 (28*/12**)	60 (40*/20**)

\* Omani participants; \*\* New Zealand participants

Sixty participants (40 from Oman and 20 from New Zealand, of which 32 were male and 28 female) took part in the experiment. They were homogeneous in that they were all undergraduates (second year) taking computer science courses. The homogeneity of the participants considered their previous learning outcomes (average GPA greater than or equal to 3.5 out of 4.5) and their existing computer programming skills. Upon completing the experiment, participants were entered into a \$50 prize draw, which aimed to externally motivate participants toward the experimental task. The unequal size of each group was due to the random way participants were assigned to conditions, carried out in the two different universities.

Forty introverted (28 Omanis and 12 New Zealanders) and 20 extraverted students (12 Omanis and 8 New Zealanders) participated. A note of the personality type considered in this study is needed here. Although there are other personality types in MBTI (i.e., sensing, thinking, or judging), we only considered extraversion (and introversion) in this study, in that the preferences for extraversion or introversion are often referred to as a first and dominant factor for learning activities in many studies (Bayne, 2004). Also, according to Bayne (2004), there is no gender issue involved in extraversion. Hence, this treatment seemed to safeguard the objective of the experiment.

### Apparatus

Two experimental systems were implemented for this study, based on a typical web-based learning system for teaching LISP, an artificial intelligence language. Figure 2 shows the non-ITS system. An intelligent tutoring system (see Figure 3) was developed based on the original ELM-ART (Brusilovsky, Schwarz, & Weber, 1996) with the permission of the developers. We used the same adaptive logic (i.e., dynamic content sequencing) with ELM-ART, changing only the content to teach LISP to provide a fair comparison with the non-ITS system shown in Figure 2.

### Procedure

Firstly, all participants in the experiment were given the MBTI test, administered by two psychologists, and then were randomly seated at a computer where either the non-ITS e-learning system or ELM-ART had been installed. On logging on to the experimental system, they were provided with simple instructions regarding the experiment. These gave information about the learning activity, the purpose of the study and the data protection policy. They then freely navigated the given system to learn the concepts of declarations, functions, and lists, and were given sufficient time to potentially learn all the material. At the end of the experiment, they were asked to answer forty questions about the content that they had learnt with the given learning system.

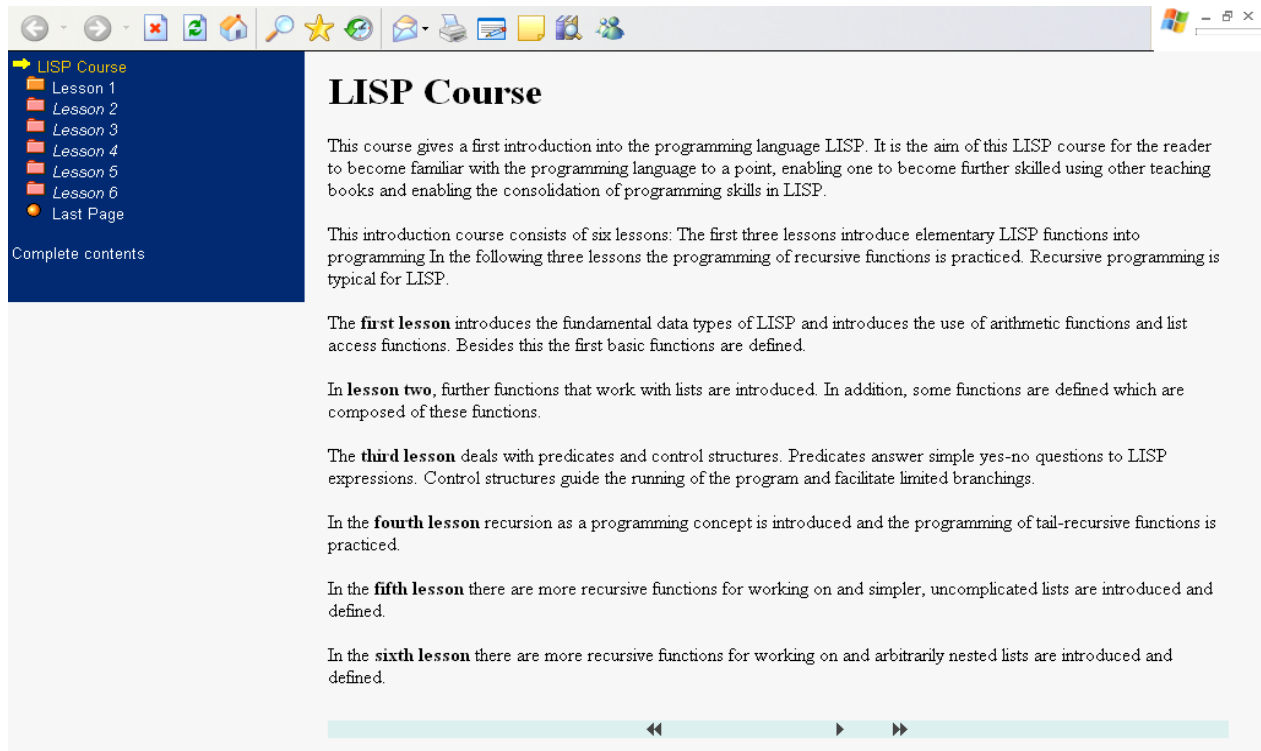


Figure 2. A web-based learning system for teaching (non-ITS system)

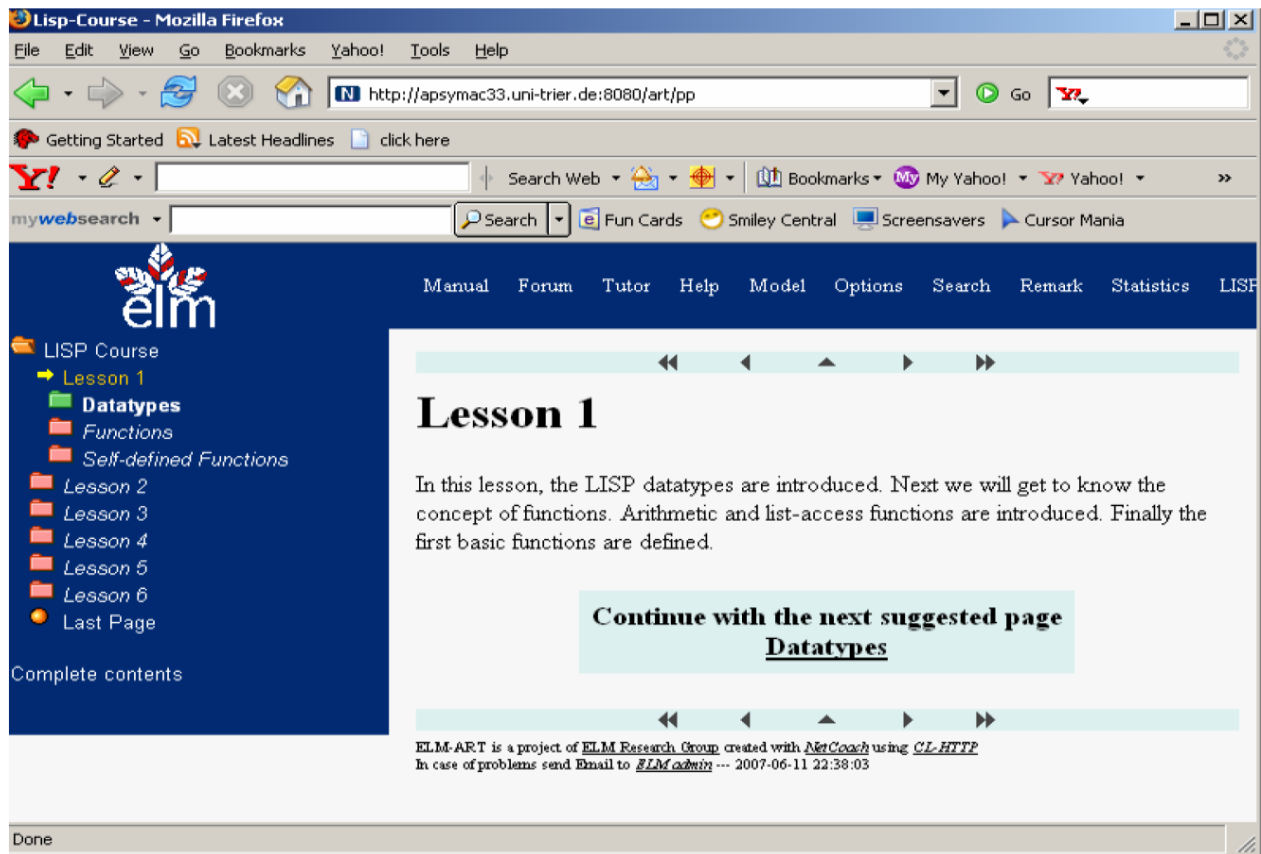


Figure 3. ELM-ART, an intelligent tutoring system, for teaching LISP (ITS system)

### Experimental design

The experimental design was a two-by-two between-subjects factorial, where personality and the computer-based learning system served as independent variables. The dependent variables were the time taken (to complete all the lessons), the number of correct answers (out of 40 test questions), the number of navigations (to check all the navigational movements), and the number of repetitions (to measure how many times learners returned to see the pages visited).

Data collected in the study were analysed using independent sample *t*-tests to compare the learning performances of the two personality types (introverted and extraverted) in the computer-based learning system.

### Results and discussion

Table 4 gives the mean performances for the learning activity. By and large, our participants seemed to derive certain benefits from ELM-ART in terms of correct answers and the number of navigations. These results clearly indicate the adaptive e-learning system could present advantages over the traditional e-learning system, which is consistent with previous studies (Calvi & De Bra, 1997).

*Table 4. A comparison of learning performances of introverted and extraverted students under the ITS and non-ITS condition*

		Extraverted		Introverted		<i>p</i>
		<i>M</i> ( <i>n</i> = 8)	<i>SD</i>	<i>M</i> ( <i>n</i> = 12)	<i>SD</i>	
ITS (ELM-ART)	Time taken (min.)	26.40	6.69	32.54	8.58	< .01
	Correct answers (%)	86.25	8.89	76.75	15.83	< .01
	no. of navigations	34.38	12.47	46.58	15.51	< .01
	no. of repetitions	5.31	3.16	13.17	9.57	< .01
		<i>M</i> ( <i>n</i> = 12)	<i>SD</i>	<i>M</i> ( <i>n</i> = 28)	<i>SD</i>	<i>p</i>
Non-ITS (Web)	Time taken (min.)	29.47	3.71	30.27	5.29	n.s
	Correct answers (%)	57.67	10.23	55.29	14.11	n.s
	no. of navigations	55.00	5.86	55.64	7.08	n.s
	no. of repetitions	12.75	3.14	12.89	3.50	n.s

Because our participants with different personality traits (introverted and extraverted) were randomly assigned to each experimental condition (ITS and non-ITS), no differences are expected in these means if there is no personality effect. The Fisher's exact test showed the association between the experimental condition and the personality was not significant (*p*-value = 0.56). Within the limits of sampling error, the effect of learning experience will thus be distributed as it is in the sampled population. Any differences that occur are therefore evidence that a particular system (either ITS or non-ITS) is sensitive to a learning style arising from a personality trait.

Comparing the learning performances of the two personality types in the non-ITS, there are no significant differences between them. But that is not the case for the ITS. Regarding the non-ITS, these results concur with earlier studies of the relationship between computer-based learning and learning styles (e.g., Felder et al., 2002; Soles & Moller, 2001) that revealed traditional e-learning systems could not allow students to pursue a multitude of different learning paths, making it very difficult to adapt to different students' learning styles. Therefore, our participants with non-ITS (12 extraverts and 28 introverts) displayed a neutral effect of their own personality traits when they were being taught by a traditional e-learning system.

As for the ITS, the extraverted students seemed to outperform the introverts in terms of the correct answers (mean 86.25 vs. 76.75) under less time taken (mean 26.40 vs. 32.54). The ITS also facilitated more efficient navigation (mean 34.38 vs. 46.58) and less repetition of the pages (mean 5.31 vs. 9.57). Though the first two measures clearly pointed out the benefits of the ITS for the extraverted students, the last two might be to some extent hard to interpret. That is, if participants took more navigations and repetitions on the course materials to learn, it can be seen that they might show more exploratory behaviour. Also, it is further countered by the fact that the lack of confidence in using

a particular system makes the users very reluctant to explore the system (Martinez, 2002). However, consider that the experiment only allowed the participants to go back to the previous pages only, thus the number of repetitions/navigations may imply an indication of how they well organise their learning experiences. Hypothetically, the user group that has a well-matched learning material structure would have efficient navigation/repetition movement. The self-organising learning activities by the introverted learners might thus show the mismatch between their personality trait and the type of their learning system, though doing more repetitions or navigations might signal higher self-directness, involvement, curiosity and so forth, in the realm of other learning contexts.

## **Conclusions**

The main conclusion to be drawn from these results is that the personality trait may have an influence on how people were able to learn with the adaptive e-learning system. This practical implication can raise awareness of the importance of personality in the design of adaptive e-learning systems, acknowledging other studies (e.g., Moallem, 2003), which concluded that learners perceive and process information in very different ways depending on their personality trait.

However, given that this study would not seek to provide a constant advantage for the extraverted in using adaptive e-learning systems, care in interpretation is needed. A comparison among the experimental conditions simply indicates the extraverted people would perform well with the personalised adaptation presented by ELM-ART, at the very least; however the introverted people tend to have more repetition (mean repetitions = 13.17 [introverted] vs. 5.31 [extraverted]) in the personalised adaptation that might cautiously suggest that focussing on the personalised adaptation should be subject to the personality trait of the learner (e.g., allowing more exploratory learning activities for the introverted). This interpretation seems to be largely in line with Moallem's findings (2003), which contend that flexible content design should be achieved by synchronising with each individual's preferred learning style.

Of course, the personality effect demonstrated in this experiment is hard to generalise to cover other learning contexts or situations in computer-based learning. It would be only possible when other systems are broadly assessed and the content and presentation styles are also tightly controlled. This possibility was therefore further explored in the subsequent study.

## **Experiment II: Content and personality effect**

Experiment II was designed to deepen the findings of Experiment I and examine if they could be applicable to other personality types, such as sensing vs. intuitive, thinking vs. feeling, and judging vs. perceiving. Also, Experiment I considered only declarative knowledge rather than procedural knowledge (e.g., knowing how to use the declarative knowledge to make an artificial intelligence programme). Hence, in building on the results of Experiment I, we further analyse preferences in learning design related to personality types while exploring procedural knowledge. In this experiment, both basic LISP primitives (declarative knowledge) and procedural LISP primitives (procedural knowledge) are addressed in examining the personality effect on learning.

These experimental manipulations are in line with several studies on personality theory (e.g., Felder et al., 2002). They generally assert that intuitive learners prefer abstract and concept-oriented approaches, so they more easily attain complex concepts and ideas, as anticipated in learning LISP, than the sensing type learner (Soles & Moller, 2001). To our knowledge, many concepts in LISP, such as lists, require understanding of complex symbolic expressions to work with data and procedures. Perhaps, this nature of LISP might credit the intuitive personality. Also, we considered evidence that the thinking personality trait facilitates more harmony with the values of logical process (Myers et al., 1998; Vincent & Ross, 2001), which is related to learning procedural LISP primitives. Finally, as to the forced learning activity (i.e., a minimum of three correct answers in a row to proceed), Myers et al. (1998) and other researchers (Felder et al., 2002; Vincent & Ross, 2001) contended that perceiving personalities prefer multiple learning choices and flexibility and dislike rigid timelines and tightly controlled interaction steps (Lawrence, 1984; 1997).



## Method

### Participants

The eighty-five participants were from Massey University. The distribution of participants across the eight groups is given in Table 5. They had some degree of homogeneity in that they were all undergraduates (third-year students) taking the computer science (CS) major at the university. Completing the experiment entitled them to be entered into three \$50 draws.

Table 5. The conditions and the sample size of each group in Experiment II

Personality	N
Extraversion (introversion)	15 (11)
Intuitive (sensing)	10 (12)
Thinking (feeling)	8 (11)
Judging (perceiving)	11(7)
Total	85

### Apparatus/procedure/experimental design

The apparatus and procedures of Experiment II were replicated those of Experiment I, except that only the ITS (i.e., ELM-ART) was used. Unlike the previous experiment, this empirical study required participants to provide a minimum of three correct answers in a row to continue to the next learning topic. This treatment was used to see if it revealed any differences by personality trait. At the end of the experiment, participants were given the extra questions of LISP. They included twenty questions of declarative and procedural knowledge of LISP (See Figure 4 and 5).

The experiment was designed as a between-subjects factorial and analysed as if it were four experiments (precise details are given in the results section.). The same dependent variables of Experiment I were used in Experiment II, namely, the time taken, the number of correct answers, the number of navigations, and the number of repetitions.

A statistical analysis was originally planned to use between-subjects analysis of variance. However, Levene's tests for heterogeneity of variance were found to be significant, suggesting that the data were not suitable for analysis of variance. For these reasons, eight Mann-Whitney U tests were adopted.

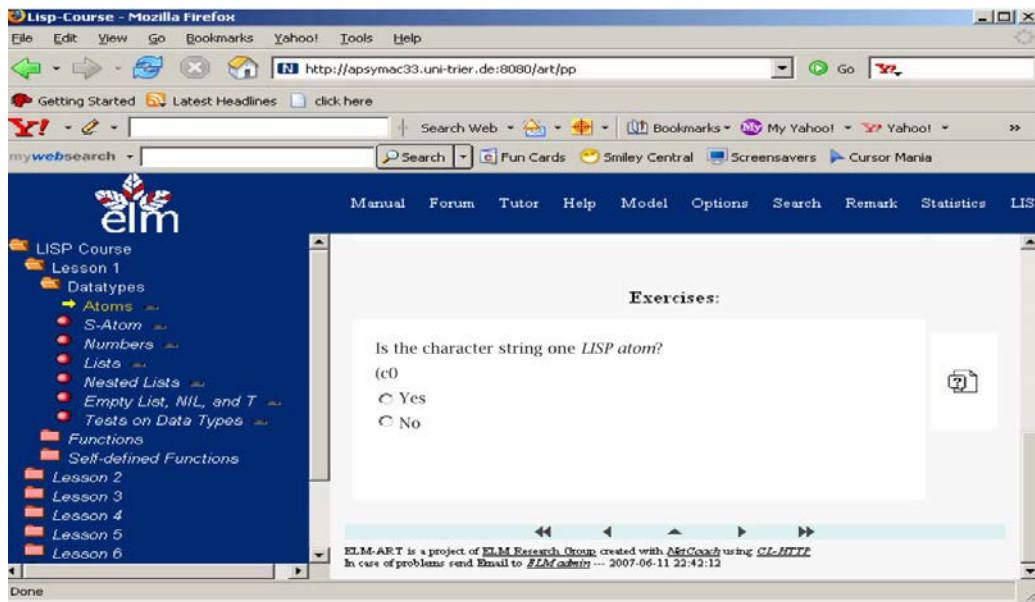


Figure 4. Example question of declarative knowledge

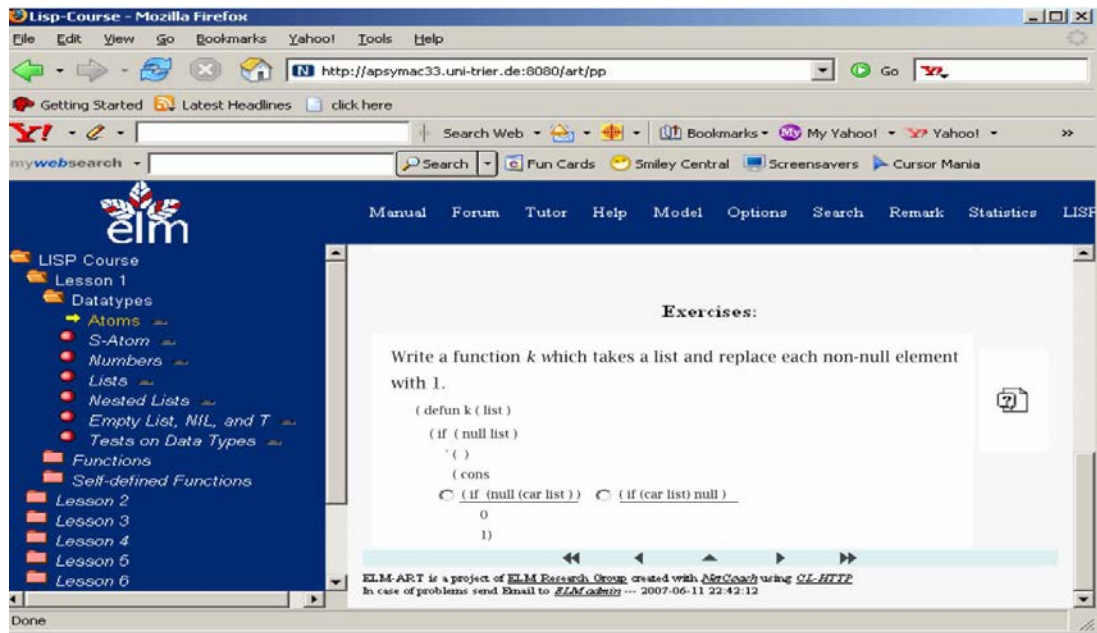


Figure 5. Example question of procedural knowledge

## Results

Table 6 gives the mean learning performance across the eight personality traits considered here. As to the procedural knowledge, the extraverted participants showed a significantly higher mean (37.11) than the introverted (28.79). Markedly different patterns were also observed within the two personality preferences (i.e., *sensing* vs. *intuitive*, and *thinking* vs. *feeling*), except the judging-perceiving pair.

Table 6. The correct answers of two types of contents (declarative vs. procedural knowledge against personality types)

	Extraversion (n = 15)		Introversion (n = 11)		U(p)	Sensing (n = 10)		Intuitive (n = 12)		U(p)
	M	SD	M	SD		M	SD	M	SD	
Declarative knowledge (20 quest.)	45.38	3.71	46.98	2.01	62.00 (n.s.)	51.77	9.09	45.83	2.06	52.00 (n.s.)
Procedural Knowledge (20 quest.)	<b>37.11</b>	<b>3.04</b>	28.79	1.24	12.00 ( $< .01$ )	27.88	4.90	<b>37.51</b>	<b>1.69</b>	8.00 ( $< .01$ )
Total score	82.49	–	75.77	–	–	79.65	–	83.34	–	–

	Thinking (n = 8)		Feeling (n = 11)		U(p)	Judging (n = 11)		Perceiving (n = 7)		U(p)
	M	SD	M	SD		M	SD	M	SD	
Declarative knowledge (20 quest.)	57.06	3.56	<b>62.98</b>	<b>7.39</b>	19.50 ( $< .05$ )	53.37	6.04	54.35	7.98	37.00 (n.s.)
Procedural Knowledge (20 quest.)	<b>24.45</b>	<b>1.42</b>	15.80	1.85	4.50 ( $< .01$ )	27.46	3.12	25.57	5.14	28.00 (n.s.)
Total score	81.51	–	78.78	–	–	80.83	–	79.92	–	–

## Conclusions and discussion

The main concern of this experiment was to see whether the results from Experiment I could be repeated. Secondly, it intended to show whether the other personality traits could affect learning performance with adaptive e-learning systems.

The first conclusion to be drawn was that, taking Experiments I and II together, the personality trait (at least extraversion-introversion) influences learning activity on adaptive e-learning systems. In this light, the work of Felder and Brent (2005) is worth noting. They contended that the introverted need to be more carefully treated to enhance their learning experience in computer-based learning.

Another empirical contribution is that our participants appear to differ in response to learning content linked to learning styles that can be derived from the personality traits. Actually, these experimental findings resemble several publications on personality theory (e.g., Felder et al., 2002; Shuck, 1999). To our knowledge, learning about lists in LISP requires an understanding of complex symbolic expressions in order to work with data and procedures. This might imply better learning outcomes for the intuitive personality (Soles & Moller, 2001). This pattern is on a par with the thinking personality trait too. Experiment II showed that they were better at learning procedural knowledge than the feeling learners, which has also been found in some other work, that the thinking personality trait values logical process (Myers et al., 1998; Vincent & Ross, 2001). It is of course not unreasonable to believe that the user may employ a different learning style, not from their assessed personality trait, in real learning situations. However, the data can be taken to suggest that care is needed when designing e-learning systems and that personality trait awareness and learning content need to be co-evolved.

## General conclusions and discussion

The main research question of this article was to investigate whether or not the learner's personality features may have certain effects on their e-learning systems, and if that is the case, how to embrace this feature in designing adaptive e-learning systems. We should note that using the personality variable to illustrate the different components of learning effect is not novel, but little has empirically done this before with adaptive e-learning systems. The personality variable was thus introduced in this study; as a result, it may conflate a number of different user modelling factors. By user modelling we mean the personality trait that can be indicative of their learning styles and preferences that can be used for structuring appropriate learning materials. This would help to design computer-based learning activity by making explicit the personality nature of the user.

The findings in our experimental contexts, regarding the likely relationship between user modelling and personality, are not new (e.g., Riding & Rayner, 1999). The results clearly showed that different learners might process the same learning material with different learning styles, so it is necessary that learning design should accord with individual differences. Hence, a practical contribution purported by this article is that understanding between the personality type and its matching computer-based learning systems design may be of central importance. It is, of course, difficult to generalise from the conditions of this study to more common computer-based learning systems that have diverse content of different domain knowledge, beyond the computer-science topics explored here. However, the data can be taken to suggest that, at the very least, care is needed when designing adaptive computer-based systems and the personality and its effect on the e-learning systems need to be considered.

No user modelling for adaptive e-learning systems can cover all design issues. This study works at the personality level of learners' preferences assumed by MBTI temperaments and thus adds what we believe to be a less-examined but crucial factor: personality. Perhaps the major limitation of the present study is that participants were only computer-science students, not from other disciplines who may have rather different personality type (Bayne, 2004). Another limitation of this study is how the personality approach will scale up to handle complex real e-learning system design projects. As mentioned in the introduction, computer-based e-learning systems should be cost-effective and easy to use by the teachers or educational providers. Considering the personality trait seems require extensive effort and this might not be as pragmatic outside the confines of a research study. With regard to the user model for adaptive e-learning system (Brusilovsky, 2001), we have yet to gain experience of the other influencing factors of learner characteristics, such as learning experience, knowledge, and prior learning process (Katuk et al, 2012). To confirm this approach as practical and exhaustive, further larger scale studies are undoubtedly needed.

## Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (20110028992; 20110021398), and Hanyang University Internal Research Fund (2012-00000001326) also helps to relieve some teaching load to finish the work.

## References

- Allen, B. L., & Kim, K.-S. (2001). Person and context in information seeking: Interaction between cognitive and task variables. *New Review of Information Behaviour Research*, 2, 1–16.
- Bayne, R. (2004). *Psychological types at work: An MBTI perspective: Psychology@Work series*. London, UK: International Thomson Business.
- Bowler, M., Bowler, J., & Phillips, B. (2009). The Big-5  $\pm$  2? The impact of cognitive complexity on the factor structure of the five-factor model. *Personality and Individual Differences*, 47(8), 979–984.
- Brusilovsky, P. (2001). Adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 11(1–2), 87–110.
- Brusilovsky, P., Schwarz, E., & Weber, G. (1996). ELM-ART: An intelligent tutoring system on World Wide Web. *Lecture Notes in Computer Science*, 1086, 261–269.
- Busato, V., Prins, F., Elshout, J., & Hamaker, C. (1999). The relation between learning styles, the Big Five personality traits, and achievement motivation in higher education. *Personality and Individual Differences*, 26(1), 129–140.
- Calvi, L., & De Bra, P. (1997). Proficiency-adapted information browsing and filtering in hypermedia educational systems. *User Modeling and User-Adapted Interaction*, 7(4), 257–277.
- Clark, R., & Mayer, R. (2008). Learning by viewing versus learning by doing: Evidence-based guidelines for principled learning environments. *Performance Improvement*, 47(9), 5–13.
- Clark, R., & Mayer, R. (2011). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. Richfield, CA: Pfeiffer.
- Costa, P., & McCrae, R. (1992). *NEO personality inventory-revised (NEO PI-R)*. Odessa, FA: Psychological Assessment Resources.
- Curilem, S.G., Barbosa, A.R., & de Azevedo, F.M. (2007). Intelligent tutoring systems: Formalization as automata and interface design using neural networks, *Computers & Education*, 49(3), 545–561.
- De Bra, P., & Stash, N. (2002). *AHA! A general-purpose tool for adaptive websites*. Retrieved August 20, 2009, from <http://www2002.org/CDROM/poster/168/>
- Deci, E., & Ryan, R. (1985). *Intrinsic motivation and self-determination in human behavior*. New York, NY: Plenum Press.
- Eklund, J., & Brusilovsky, P. (1998, September). *Interbook: An adaptive tutoring system*. Paper presented at the Apple University Consortium Conference: Flexible Learning: Exploring the Myths and Realities, Melbourne, Australia.
- Entwistle, N., & Entwistle, D. (1970). The relationships between personality, study methods, and academic performance. *British Journal of Educational Psychology*, 40(2), 132–143.
- Felder, R., & Brent, R. (2005). Understanding student differences. *Journal of Engineering Education*, 94(1), 57–72.
- Felder, R., Felder, G., & Dietz, E. (2002). The effects of personality type on engineering student performance and attitudes. *Journal of Engineering Education*, 9(1), 3–17.
- Furnham, A. (1992). Personality and learning style: A study of three instruments. *Personality and Individual Differences*, 13(4), 429–438.
- Furnham, A., & Medhurst, S. (1995). Personality correlates of academic seminar behaviour. *Personality and Individual Differences*, 19(2), 197–208.
- Gardner, H. (1983). *Frames of Mind*. New York, NY: Basic Books.
- Graesser, A., Chipman, P., Haynes, B., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education*, 48(4), 612–618.

- Heaven, P. (1989). Attitudinal and personality correlates of achievement motivation among high school students. *Personality and Individual Differences*, 11(7), 705–710.
- James, L., & McIntyre, M. (2000). *Conditional reasoning test of aggression: Test manual*. Knoxville, TN: Innovative Assessment Technology.
- Katuk, N., Kim, J. E., & Ryu, H. (2012). Experience beyond knowledge: Pragmatic e-learning systems design with learning experience. *Computers and Human Behavior*, 29(3), 747–758.
- Khan, A., & Radcliffe, K. (2005). *Mind shapes: Understanding the differences in thinking and communication*. St. Paul, MN: Paragon House.
- Kogan, N. (1971). Educational implications of cognitive styles. In G. S. Lesser (Ed.), *Psychology and Educational Practice* (pp. 242–292). Chicago, IL: Scott Foresman.
- Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Englewood Cliffs, NJ: Prentice Hall.
- Lawrence, G. (1984). A synthesis of learning style research involving the MBTI. *Psychological Type*, 8(2), 2–15.
- Lawrence, G. (1997). *Looking at type and learning styles*. Gainesville, FL: Center for Applications of Psychological Type.
- Martinez, M. (2002). Beyond classroom solutions: New design perspectives for online learning excellence. *Educational Technology & Society* 5(2), 1–6.
- McCrae, R., & Costa, P. (1989). Reinterpreting the Myers-Briggs type indicator from the perspective of the five-factor model of personality. *Journal of Personality*, 57(1), 17–40.
- McCrae, R. (2002). NEO-PI-R data from 36 cultures: Further intercultural comparisons. In R. McCrae & J. Allik (Eds.), *The Five-Factor Model across Cultures* (pp. 105–126). New York, NY: Kluwer Academic/Plenum Publishers.
- Messick, S. (1976). *Personality consistency in cognition and creativity, individuality in learning*. San Francisco, CA: Jossey-Bass.
- Millon, T. (1990). The disorders of personality. In L.A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 339–370). New York, NY: Guilford.
- Moallem, M. (2003). Applying learning styles in an online course. *Academic Exchange Quarterly*, 7(4), 209–215.
- Moundridou, M., & Virvou, M. (2003). Analysis and design of a web-based authoring tool generating intelligent tutoring systems. *Computers & Education*, 40(2), 157–181.
- Myers, I., McCaulley, M., Quenk, N., & Hammer, A. (1998). *MBTI Manual* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.
- Pulford, B., & Sohal, H. (2006). The influence of personality on HE students' confidence in their academic abilities. *Personality and Individual Differences*, 41(8), 1409–1419.
- Riding, R., & Rayner, S. (1999). Cognitive styles and learning strategies: Understanding differences in learning and behavior. *Educational Technology*, 15(3), 222–241.
- Sampson, D., Karagiannidis, C., & Kinshuk (2002). Personalised learning: Educational, technological and standardisation perspective. *Interactive Educational Multimedia*, 4, 24–39.
- Shuck, A. (1999). Assessing pharmacy students' learning styles and personality types. *American Journal of Pharmaceutical Education*, 63(10), 27–33.
- Smith, E., & Riding, R. (1999). Cognitive style and instructional preferences. *Instructional Science*, 27(5), 355–371.
- Soles, C., & Moller, L. (2001). Myers-briggs type preferences in distance learning education. *International Journal of Educational Technology*, 2(2), 160–167.
- Stipek, D. (2001). *Motivation to learn: Integrating theory and practice*. Boston, MA: Allyn and Bacon.
- Vincent, A., & Ross, D. (2001). Personalize training: Determine learning styles, personality types and multiple intelligences online. *The Learning Organization*, 8(1), 36–43.
- Weber, G., Kuhl, H., & Weibelzahl, S. (2002). Developing adaptive Internet based courses with the authoring system NetCoach. *Lecture Notes in Computer Science*, 2266, 226–238.
- Zhang, L. (2003). Does the big five predict learning approaches? *Personality and Individual Differences*, 34(8), 1431–1446.

## Leveraging Quiz-based Multiple-prize Web Tournaments for Reinforcing Routine Mathematical Skills

**Ana I. González-Tablas\*, José M. de Fuentes, Jorge L. Hernández-Ardieta and Benjamín Ramos**

Computer Science and Engineering Department, Universidad Carlos III de Madrid (SPAIN)  
Avda. de la Universidad, 30; 28911, Leganés (SPAIN) // aigonzal@inf.uc3m.es // jlfuentes@inf.uc3m.es // jlhernan@inf.uc3m.es // benja1@inf.uc3m.es

\*Corresponding author

(Submitted October 04, 2011; Revised June 06, 2012; Accepted July 25, 2012)

### ABSTRACT

In Higher Education Engineering studies, there exists the need of engaging students in performing drill and practice activities with the goal of reinforcing routine mathematical skills. The usual optionality of these tasks entails the risk of students not fulfilling them in an effective way. Although competitive approaches are not a trend in current educational practices, there are several arguments that support their use in higher education for addressing this problem, provided that the negative effects of pure interpersonal competitions are mitigated. Recent works in this direction have obtained mixed results. In this work, results of a quasi-experiment conducted to evaluate the effects on students' outcomes and satisfaction of a competitive web-based drill game are presented. Moreover, a novel multiple-prize soft competition design that considers several stages combined with public ranks is proposed, with the goal of minimizing the disadvantages of pure interpersonal competitions. Results indicate that the tournament is highly effective in increasing students' outcomes and that students' satisfaction is rather high.

### Keywords

Competitive learning, Higher education, Routine mathematical skills, Multiple-prize tournaments, Soft competitions

### Introduction

As a result of the Bologna Process, students' mean workload in European Higher Education is 50 hours per week. Approximately half of this workload corresponds to teacher-assisted learning activities and the other half to teacher-unassisted ones. The Bologna Process promotes learning activities to be student-centered (Conference of European Ministers Responsible for Higher Education, 2009) because these activities are usually more engaging for students, who commonly spend more time and effort on them, and facilitate a higher quality learning. On the contrary, non-student-centered activities are usually less motivating for students, with a high risk of students not fulfilling the activity or at least not putting enough effort on it.

However, there are some non-student-centered activities that are simply to deploy and that are effective in helping students achieve their learning goals. This is the case of homework drills, which are traditionally used to reinforce routine mathematical skills in higher education. Acquiring routine mathematical skills is very important as they underlie the whole process of learning to think mathematically (Schoenfeld, 1992). First, some sessions are devoted to explaining the targeted theoretical concepts and illustrate them with examples, and, second, some reinforcement sessions take place. During the reinforcement sessions, students solve a set of proposed exercises and the teacher—or the students assisted by the teacher—provides a detailed solution to all or part of the exercises using a blackboard or an electronic slide projector. As the exercises address routine mathematical skills, it is expected that the skills will be reinforced if the students repeat the process by themselves several times. This is usually achieved by an unassisted activity: providing a homework sheet and later publishing the solutions to the exercises.

Because doing homework sheets of this type is usually both time consuming and not very motivating for students, there is a high risk of students' not completing the activity properly. The main consequence is that, afterwards, they are not prepared to successfully handle the rest of the course goals. Even if homework drills are made mandatory for passing the course and students' submissions get assessed, it is highly improbable that students' motivation increases. Contrarily, students' plagiarism and teachers' workload are expected to grow, being the last of the two increments, to the detriment of the teachers' dedication to scaffolding students in more challenging learning activities. In the described scenario, there exists the need of engaging students in learning a particular basic skill.

A well-known pedagogy of engagement in higher education is cooperative learning. As defined by Smith, Sheppard, Johnson, & Johnson (2005), “cooperative learning is the instructional use of small groups so that students work together to maximize their own and each other’s learning.” Competitive learning is an alternative student engagement pedagogy with respect to individualistic learning. In competitive learning, according to Johnson, Johnson, & Stanne (1985), “when one student achieves his or her goal, all others with whom he or she is competitively linked fail to achieve their goals.” Contrarily, individualistic learning is defined in Johnson et al., (1985) as those situations where “the goal achievement of one student is unrelated to the goal achievement of others”.

Smith et al. (2005) summarize the main results of the meta-analysis made on research conducted between 1924 and 1997 on cooperative, competitive, and individualistic learning for higher education. Analyzed works provide clear evidence on the superiority of cooperative learning. Some researchers also highlight the negative effects on achievement motivation that competitive learning approaches may have in pre-higher education students (Johnson et al., 1985; Lam, Yim, Law, & Cheung 2004). As a result, although not all results support these theories (e.g., Rzoska & Ward, 1991), nowadays, pure interpersonal competitive activities are not a trend in educational practices.

However, several arguments support the use of competitive learning activities in higher education for reinforcing routine mathematical skills. On one hand, competition and performance comparison with peers (ranks) are very powerful motivating factors to achieve better performance in adult settings, even in the absence of rewards (Blanes i Vidal & Nossol, 2011). On the other hand, recent research in the context of pre-higher education has proven that competitive activities (taking the form of a game) can be as effective as cooperative ones in improving student outcomes, and that students prefer the learning activities that present some type of social competition (students against students or teams against teams) (Ke & Grabowski, 2007; Kebritchi, Hirumi, & Bai, 2010). Moreover, in the context of economics, researchers are proposing new prize distribution structures that lessen the negative effects of pure interpersonal competitions (Moldovanu & Sela, 2001; Cason, Masters, & Sheremeta, 2010).

In the context of higher education, few research works have studied, with different results, the effects that competitive drill games have on student outcomes and attitudes (Cameron & Dwyer, 2005; Ebner & Holzinger, 2007; Regueras, Verdú, Muñoz, Pérez, de Castro, & Verdú, 2009). Cameron and Dwyer (2005) analyze the effects of using an interpersonal competitive game with several feedback modes (no feedback, knowledge of response feedback, and elaborative feedback) in comparison to not using it. The games’ quizzes aim to teach human heart concepts. Students compete to have the higher score and the fastest time to correctly answer the questions. Authors conclude that competition by itself, which they expected to achieve increased student outcomes compared to those obtained when the game was not used, does not have a significant effect except when it is accompanied by a feedback. Ebner and Holzinger (2007) also analyze the effects of using an interpersonal competitive quiz-based game, the Internal Force Master, with civil engineering students. Authors’ main hypothesis is that by playing in a competitive environment, students will learn more than non-playing students because incidental learning will happen. However, results of the experiment show no significant differences between playing and non-playing students’ outcomes.

On the other hand, Regueras et al. (2009) obtain very significant positive results when they use the QUEST (Quest Environment for Self-managed Training) system to teach Communication Networks to students in the Telecommunication Engineering program. QUEST allows students to compete in teams to solve different types of challenges. Results show that grades of playing students are higher compared to non-playing ones, but that grades are not correlated to game scores. Students’ satisfaction is also found to be positive in spite of the competitive characteristic of the game.

As can be seen from the literature review (see also Figure 1 for a detailed comparison), recent experiments analyzing the effects of competitive educative games designed to make higher education students perform drills have obtained mixed results. Further experiments in this direction should be a valuable contribution in order to clarify this issue.

The goal of this work is to assess the effects on student outcomes and satisfaction that a competitive web-based drill game may have when it is used for reinforcing routine mathematical skills of higher education students instead of the traditional homework drills. The game is designed as a quiz-based tournament with a multiple-prize structure and public score ranks. The tournament is designed to lessen the disadvantages of pure interpersonal competitions and motivate students’ participation. Because it is implemented using computer-based technologies, a computer-based trainer is also introduced in the instruction to minimize the differences with lecture-based approaches. We expect that students who participate in the tournament and use the trainer achieve higher outcomes than those obtained by those



of the control group. We also expect that students perceive the experience as more satisfactory than the lecture-based approach. The use of the trainer (only) is also expected to benefit student outcomes and satisfaction, but to a lesser extent.

Competitive approaches in higher education				
	(Cameron and Dwyer, 2005)	(Ebner and Holzinger, 2007)	Our proposal	(Regueras et al., 2009)
Population size	422 students	141 students	133 students	200 students
Instruments	"Heart Attack" (web-based game) + tests	"Internal Force Master" (web-based game) + tests + questionnaires	Quiz-based multiple-prize tournament (web-based game) + tests + questionnaires	"QUEST" (web-based game) + tests + questionnaires
Type of problem	Quiz (multiple-choice question)	Quiz (multiple-choice question – selection of drawing depicting correct internal forces of a static determinate system)	Quiz (multiple-choice and numeric answer questions)	Variable (usual types in current assessment: from short questions to complex problems)
Area	Medicine – Parts and functions of human heart	Civil engineering – Theory of Structures	Computer engineering – Cryptography	Communication engineering – Data communication networks
Type of game	Quiz (drill-and-practice)	Quiz (drill-and-practice)	Quiz (drill-and-practice)	Challenges (proposed by students or teachers) and quiz
Quiz feedback	Three types of feedback: (1) no feedback, (2) knowledge response feedback and (3) elaborative feedback	Knowledge of response feedback	Knowledge of response feedback	Knowledge of response feedback according to a set of previously known criteria
Assesment (evaluation)	Automatically assessed by the system	Automatically assessed by the system	Automatically assessed by the system	Assessed by the author of the question
Competition type	Interpersonal	Interpersonal	Interpersonal	Intergroup (groups of two students)
Soft competition	No	No	Yes	Yes
Competition in several steps	Single contest (single step)	Single contest (but several steps)	Tournament with multiple contests	Single contest (but several steps)
Existence of rank order	Yes	Yes	Yes	Yes – both individual and team rankings
Rank order public/private	Public (score and time)	Public (higher scores and best of the month)	Public (score)	Public (score)
Prize structure / Subprizes	– (no prize is mentioned)	– (no prize is mentioned)	Multiple prizes. Contest prizes: first five of each contest receive a reward. Tournament prizes: rewards are assigned according to the tournament rank position and a set of quota-limited fixed prize categories	Multiple prizes. Rewards count a maximum of 40 % of laboratory mark (proportional part respect the best student)
Duration	1 playing session of 20 minutes	Game available to students during 3 months	Tournament took place during 5 days	QUEST was used along the 5 months of the course
Experiment design	(C) Non-playing vs (T1) playing with no feedback vs (T2) playing with knowledge of response feedback vs (T3) playing with elaborative feedback	(C) Non-playing vs (T1) playing	(C) Non-playing vs (T1) quiz-based trainer (with elaborative feedback) vs (T2) quiz-based trainer (with elaborative feedback) followed by playing (with knowledge of response feedback)	(C) Non-playing vs (T1) playing
Results on outcomes	~ (Note 1) Significant differences exist in the following treatment comparisons respect to the specified aspect: Drawing (T3 > C, T3 > T1), Identification (T3 > C, T3 > T1), Terminology (T3 > C, T3 > T1), Comprehension (T3 > C), Composite score (T2>C, T3 > C, T3 > T1)	X No significant difference exist between C and T1 outcomes	✓ Significant differences exist between T2 > C and T2 > T1. No significant differences exist between T1 and C	✓ Significant difference exist between T1 > C
Results on motivation/satisfaction	– (it is not analyzed)	Enjoyability factor moderately high (between 1 and 2 being [1] "I enjoy the game" and [5] "I do not enjoy the game")	Students' overall satisfaction is moderately high (mean of 22.38 over a scale of [6] for the lowest overall satisfaction and [30] for the highest overall satisfaction)	Students are satisfied with the use of the QUEST tool (average score of 36.5)
Note 1: Results seem more dependent on the type of feedback than on playing the competitive game in the same conditions than the other settings (knowledge of response feedback)				

Figure 1. Detailed comparison of related work with our proposal

## Methodology

### Educational context

The Cryptography and Computer Security (CCS) course is mandatory in the curriculum of the Grade on Informatics Engineering at the University Carlos III of Madrid (Spain). The CCS course takes place in the second semester of the second year and addresses the fundamentals of cryptography and computer security. It comprises 32 sessions of 90 minutes distributed in 14 weeks. Ninety-four percent of the course's time is devoted to cryptography (56% to theory,



25% to exercises, and 13% to labs), being covered all the topics of the unit AL/CryptographicAlgorithms included in (CS2008 Interim Review Task Force of the ACM and IEEE Computer Society 2008). The remaining 6% covers the basic concepts of unit PF/FoundationsInformationSecurity.

A large part of the course's problems is based on basic modular arithmetic. However, very few sessions can be devoted to them as they are not the main focus of the course. Two sessions are devoted to explain theoretical concepts and illustrate them with examples. A third session reinforces the students' skills with drills. Each of the three sessions is performed in a lecture-based mode, using an electronic slide projector and blackboard. Optional exercises are also provided through the course's web page as homework to complete the reinforcement.

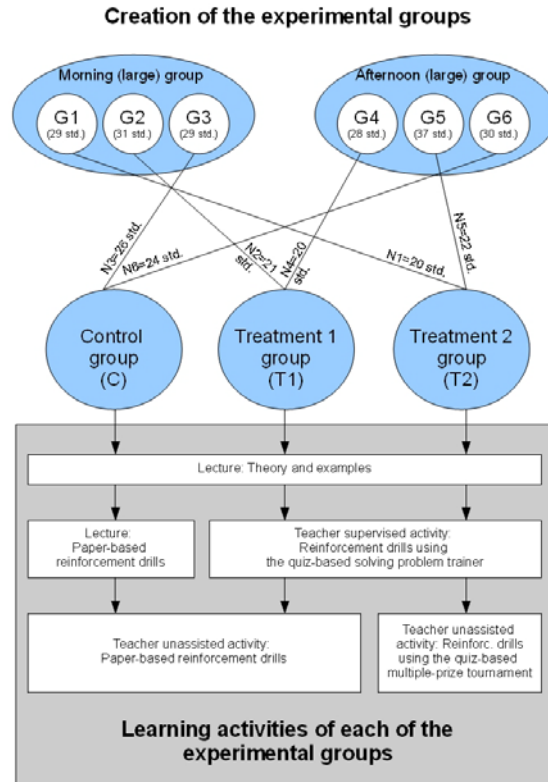


Figure 2. Creation of the experimental groups and learning activities assigned to each of them

### Design of the experiment and participants

The approach taken in this work consists in using a quiz-based multiple-prize tournament instead of the paper-based homework drills. In the experiment, the lecture-based approach of the third session is also substituted by a quiz-based problem solving trainer that guides the students in solving the exercises at their own pace. Because the use of the trainer may also affect student outcomes and satisfaction, two incremental treatments are analyzed respect to a control group (see Figure 2): the use of the trainer (treatment 1), and the use of the trainer followed by the participation in the tournament (treatment 2). Students assigned to the control treatment receive lecture-based instruction as described in the section Educational Context.

The subjects participating in the study are 133 students ( $N_1 = 20$ ,  $N_2 = 21$ ,  $N_3 = 26$ ,  $N_4 = 20$ ,  $N_5 = 22$ , and  $N_6 = 24$ ) enrolled in six small groups of the spring 2011 CCS course. Groups 1, 2 and 3 are aggregated for the magisterial lessons in one classroom (morning large group), and the same happens with groups 4, 5 and 6 (afternoon large group). Each small group of the large groups is randomly assigned to one experimental group (see Figure 2). As a result, the control (C), treatment 1 (T1) and treatment 2 (T2) groups have respectively 50, 41, and 42 students. The analysis has considered only those enrolled students who have followed the continuous assessment modality, have participated significantly in the experiment, and for whom all grades used in the study were available.

## Materials

The trainer system consists of five web-based quizzes that guide students in solving five detailed exercises in a step-by-step mode. The exercises seek to reinforce the skills of solving equations of the type " $a \cdot x = b \text{ mod. } n$ ," that is, if it has any solution (single or multiple), and uses several methods to calculate the multiplicative inverse of  $a$  modulo  $n$ . Immediate elaborative feedback is provided and multiple attempts are allowed for each question. The trainer system is integrated in the university's web-based learning platform, which builds on Moodle (<http://moodle.org>). Figure 3 presents two sample screen shots where details on the type of elaborative feedback are provided.

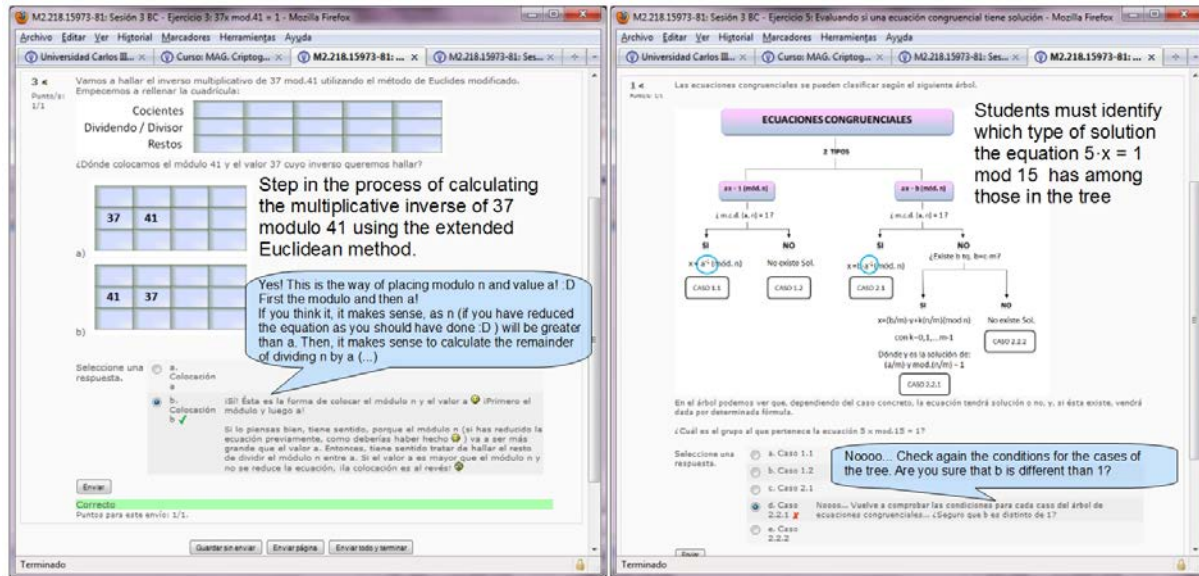


Figure 3. Sample screen shots of the trainer system

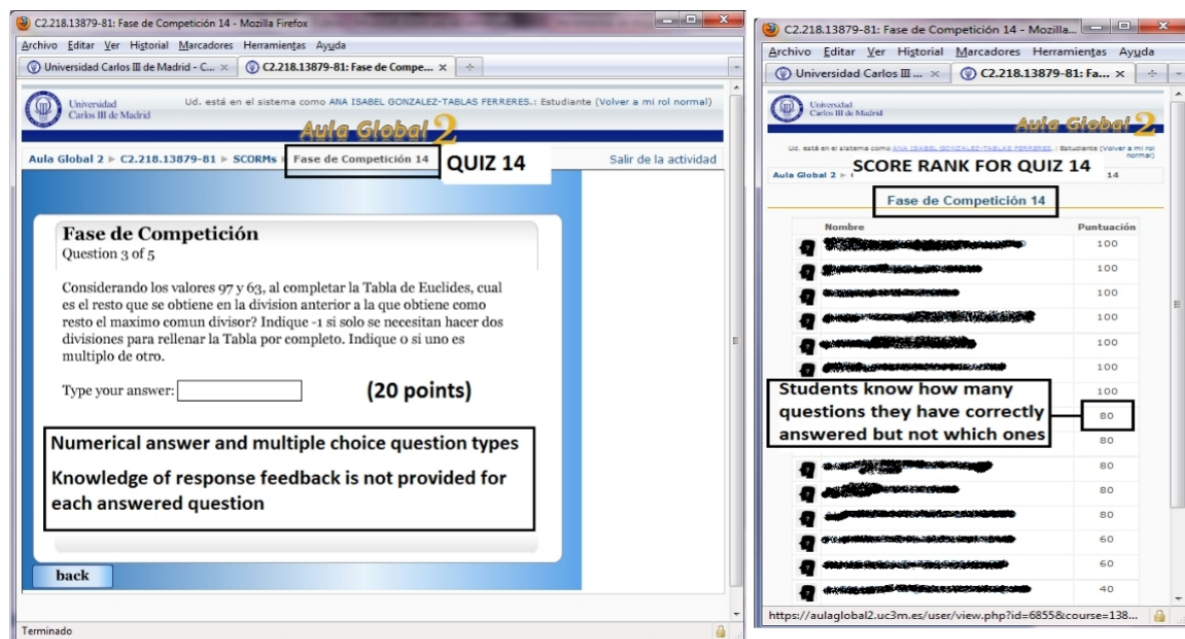


Figure 4. Sample screen shots of the tournament system

The tournament system consists of five contests, each of them posing to the students a set of five web-based quizzes. Each quiz comprises five questions (see an example in Figure 4, left). Each quiz has associated a public score rank (built as a PHP web page) that students can consult at any time (see Figure 4, right). Students may attempt to answer

each quiz multiple times while its contest is open (around 24 hours). Although multiple attempts are allowed, students are strongly encouraged to make only two. Note that knowledge of response feedback is not directly provided but students know how many questions of the quiz they have correctly answered by looking it up in the quiz score rank.

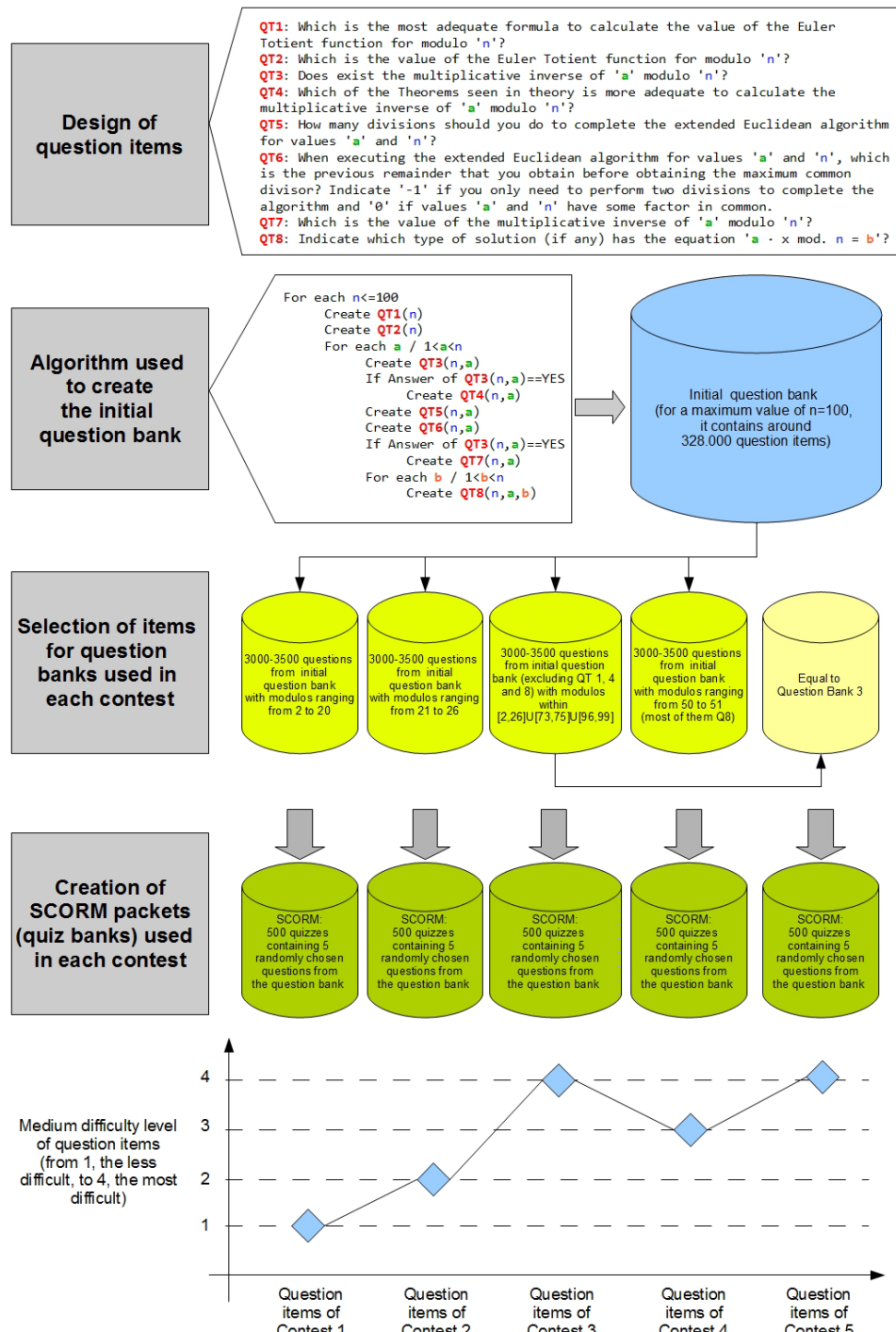


Figure 5. Details on the question items and on the construction of the quiz banks

To build the quiz banks, first, an initial question bank was created containing eight types of questions on the general equation  $a \cdot x = b \text{ mod. } n$  (see Figure 5). They were designed with the trainer questions as a model. Most of the questions are such that can be automatically generated and assessed, but not automatically answered by the students

(to force them performing some calculations). A Java program was developed to generate the initial question bank. From it, for performance reasons, four question banks (light green in Figure 5) were extracted containing 3000–3500 questions with different levels of difficulty (see Figure 5). Each question bank was associated to one of the contests (third question bank was reused in contest five). From the question banks, five quiz banks (darker green in Figure 5) were built containing 500 different quizzes, each quiz comprising five questions randomly selected from the question bank. The quiz banks were implemented as sharable content object reference model (SCORM) packets. For each contest, five instances of the SCORM packet associated to the contest were published in the university's internal course web page. Each time a student attempts to answer a quiz, the SCORM package presents a (probably) different quiz by selecting it randomly from the quiz bank.

The paper-based reinforcement drills provided as homework just contained a set of four equations of the type  $a \cdot x = b \text{ mod. } n$  that students have to solve.

### Evaluation algorithm and tournament prize structure

Each question of the tournament has a value of 20 (game) points if correctly answered. The higher score obtained by a student among all attempts made on a quiz is the one registered by the system and the one publicly available in the quiz score ranks. However, the tournament prize structure builds not only on the quiz scores but also on the number of attempts used by a student on that quiz. Using both data sets, we calculated weighted quiz score as follows: if the number of attempts is greater than four, 20 points are subtracted from the student's quiz score; if greater than eight, 40 points, etc. Figure 6, which illustrates (with real data) the tournament prize calculation, shows on the left several examples of a weighted quiz score calculation. For example, for the quiz number four within contest number three, the student has obtained 80 points (four questions correctly answered) and used seven attempts. Therefore, the weighted quiz score is calculated 20 points less than the original quiz score.

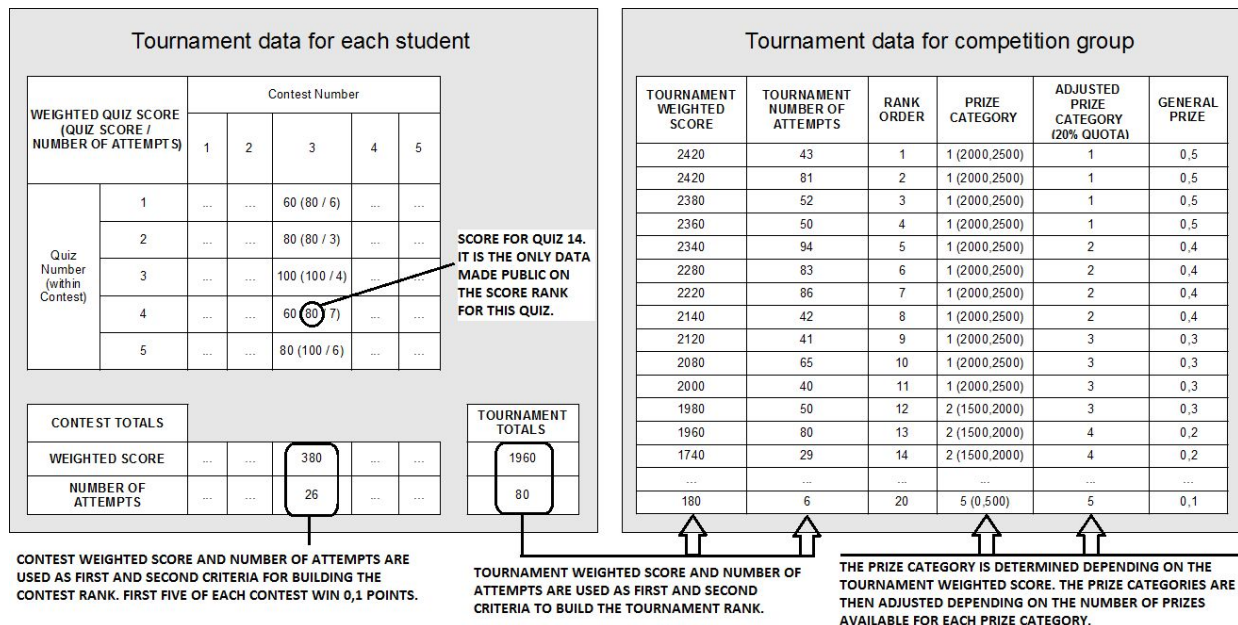


Figure 6. Illustration of the tournament prize calculation

Then, contest totals (weighted score and number of attempts) are calculated as the sum of their values for each of the five quizzes of a contest. In the left lower part of Figure 6, totals of contest 3 are calculated for one student. Afterwards, contest ranks (not shown in the figure) are built using as criteria the contest totals (first criterion, contest weighted score; and second criterion, contest number of attempts). The first five students in each contest rank win a reward of .1 (real) points. At the end of the course, the (real) points won in the tournament are added to the student's grade only if the student has passed the course.

After all the contests have taken place, tournament totals (tournament weighted score and tournament number of attempts) are calculated for each student as the sum of their corresponding values for the five contests. A tournament rank is built using, as first criterion, the tournament weighted score and, as second criterion, the tournament number of attempts (see right part of Figure 6). Depending on the position of a student in the tournament rank, the student wins one among five possible rewards or general prizes (.5, .4, .3, .2, and .1 real points) according to the algorithm described next.

First, a prize category is assigned to each student depending on the value of the student's tournament weighted score: first prize category for scores falling within (2000, 2500), second for (1500, 2000), and so on. However, the number of students that may obtain the same prize is restricted by quota (to avoid collusion between the students). Then, the prize category assigned to each student may be adjusted according to this quota. In this experiment, the quota of each category was fixed to the 20% of the number of students participating in the tournament. In Figure 6, 20 students participated in the tournament, so only four students can be assigned to the same category. Eleven students are assigned initially to the first prize category. After adjusting them, only the first four keep the first category, while the second four are assigned to the second one, and the eleventh student is assigned to the third. The students who were initially assigned to the second prize category are displaced to lower prize categories. Figure 7 shows a more formal description of the algorithm used to determine the students' rewards.

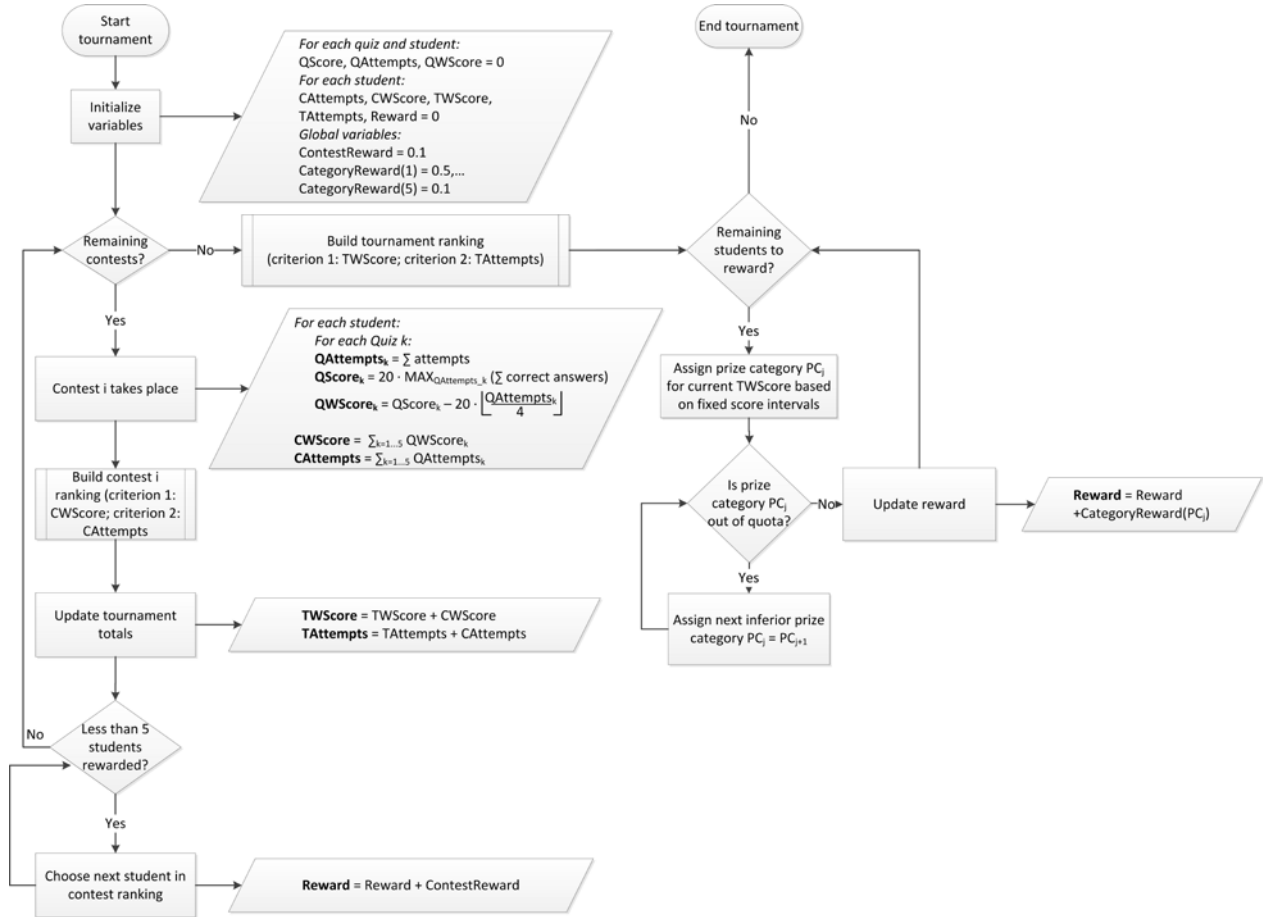


Figure 7. Illustration of the algorithm that determines the tournament rewards

As far as authors know, the tournament prize structure used in this work is novel in educative settings although its features are similar to some of the structures used in current sports competitions (such as the rank-based multiple-prize structures with quota-limited prize categories found in some fencing competitions or having both contest rewards and tournament rewards of some cycling competitions). The design of the tournament prize structure is the result of applying the following reasoning:



- A multiple-prize structure based on prize categories associated to fixed performance ranges has been chosen in order to promote participation in the tournament (there are as many prizes as participants). Proportional prizes have been avoided as our goal was that students achieved a certain level of performance.
- Quota-limited prize categories have been established to avoid students' collusion (all falling in the same prize category) and promote individual reinforcement effort.
- Several contests take place consecutively to promote skills reinforcement in time (five days) and contest rewards promote excellence in each contest.

## Instruments

In order to assess the homogeneity of the control and experimental groups, we designed an initial level test containing five multiple choice questions on basic mathematic concepts for undergraduate engineering students (inverse of a matrix, maximum common divisor, inequality system, binary to decimal transformation, and multiplicative inverse modulo  $n$ ). The test was performed under supervision, and student participation was heterogeneous and not high enough (65%, 47%, and 60% for groups C, T1, and T2, respectively). Note that attendance and assignments are optional, and that only exams are mandatory. Therefore, as it was not adequate to use the results of this test, the grades of a previous (teacher-unassisted) course assignment were used to assess the groups' homogeneity. It consisted in analyzing a real security case (students, in pairs, must seek and analyze a recent security incident with impact in the media). The maximum grade for this assignment was 1 point.

The same drills were prepared in paper and electronic formats for the third teacher-assisted session devoted to modular arithmetic. The paper format just consisted in a list of equations (of the type  $a \cdot x = b \text{ mod. } n$ ) to solve and no data was collected regarding them. The electronic format was in fact the trainer system, which just collected the participation of the students in its quizzes and their results.

Regarding the teacher-unassisted reinforcement drills, the tournament system registered the students that have accessed to its quizzes, the higher score obtained by each student for each quiz, and the number of attempts used for it. Data were extracted from the university's learning platform and processed with Microsoft Excel software and a Ruby program developed specifically for calculating each student's prize.

An exam on modular arithmetic was elaborated in both paper and electronic formats (with the same questions) to measure the effects of the treatments on students' outcomes. The exam contained two exercises (the resolution of two equations of the type  $a \cdot x = b \text{ mod. } n$ , one with one solution and another with multiple solutions). The exam was performed under supervision and the maximum grade was .5 points.

Finally, a survey was designed to measure the students' satisfaction after experiencing treatments 1 and 2. The survey consisted of 6 items about the following aspects when compared to lecture-based methods: (1) perceived effectiveness, (2) ease of use, (3) created motivation, (4) experienced fun, (5) preference for the treatment method, and (6) suggestion of reuse in other modules of the course. The answers followed a five-point Likert-type scale (from "Strongly Disagree" to "Strongly Agree"). The satisfaction was calculated as the sum of the values of the six items.

## Procedure

The procedures applied to each experimental group are depicted in Figure 8. The assignment analysis of a real security case was explained in the session four of the course (week two) and students handed in the report during week three. The first two sessions devoted to modular arithmetic (sessions five and six) took place during week three and followed a lecture-based approach. In the eighth session of the course, the third session devoted to the target topic takes place. In this session, students of group C were instructed using a lecture-based approach while groups T1 and T2 used the trainer system supervised by a teacher. In this session, students of group T2 were also briefly instructed on the tournament. A different tournament was launched for each of the small groups assigned to T2. Each contest took place in one of the five days following session eight, being open a full day and then closed. Participation in the tournament was voluntary and students decided the quizzes that they will answer. The exam took place a week after session eight. Group C took it in paper format, and groups T1 and T2 did it through the web. Finally, students of groups T1 and T2 were asked to voluntarily answer the satisfaction survey.

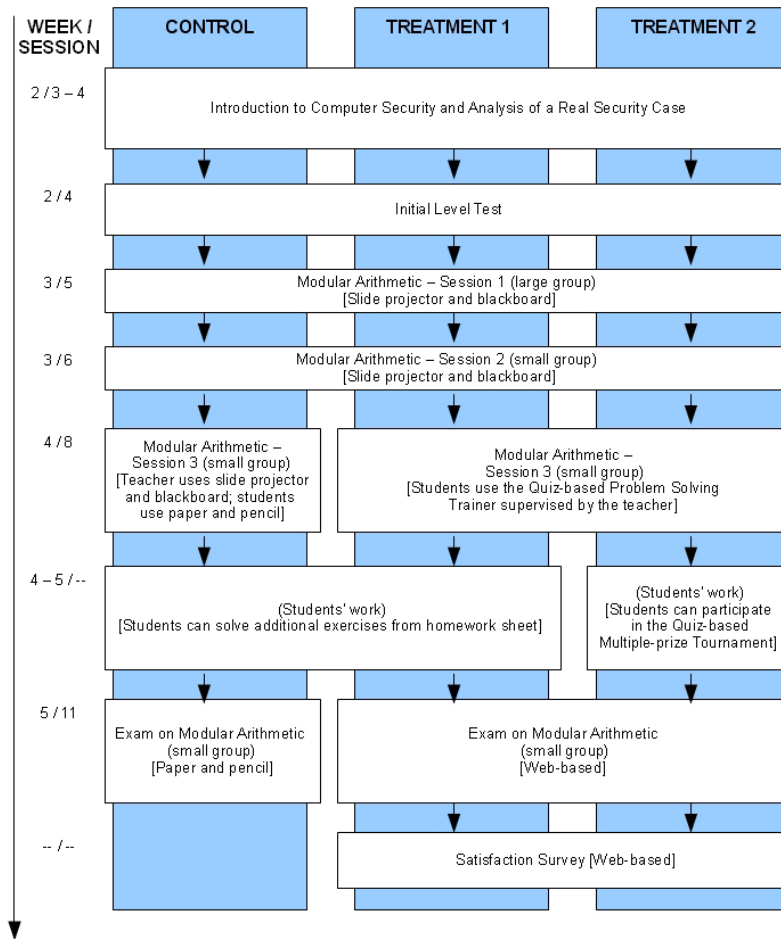


Figure 8. Experimental procedures applied to control, treatment 1 and treatment 2 groups

## Data analysis

The R environment for statistical computing (<http://www.r-project.org/>) has been used to analyze the collected data. First, the students' grades for the real security case analysis have been analyzed to evaluate the homogeneity of the different experimental groups. Second, the students' grades for the modular arithmetic exam have been analyzed to evaluate the effects of each treatment on students' performance. For both analyses the following tests have been conducted: the Saphiro-Wilk test for testing normality, the Kruskal-Wallis and Fligner-Killeen tests for testing the existence of differences between the location parameters of the grades distributions for all groups, and the Mann-Whitney test for testing the differences between the means of the grades for each pair of groups. The significance level has been fixed at  $\alpha = .05$  for all tests, except for the Wilcoxon-Mann-Whitney test, where the Bonferroni correction was applied as three comparisons are made with the same data ( $\beta = .01667$ ). Besides, the Pearson's correlation test was also performed to analyze the possible correlation between the students' outcomes and the students' tournament scores. Note that it is assumed that the grades come from a continuous domain although they have been discretized. For the analysis of the students' satisfaction, the Kruskal-Wallis and Fligner-Killeen tests have again been used.

## Results and discussion

The distributions of the normalized grades are summarized in the box plots shown in Figures 9 and 10. The Shapiro-Wilk test for all data sets indicates that data is not normally distributed, forcing the use of the mentioned non-parametric tests to analyze the effects of the treatments.

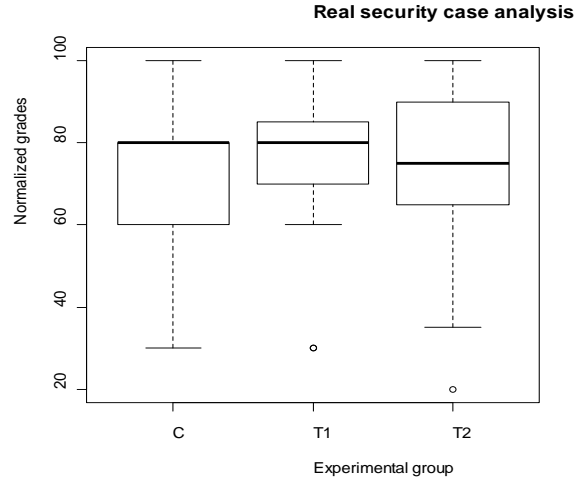


Figure 9. Box plot of the normalized grades obtained in each experimental groups for the real security case analysis

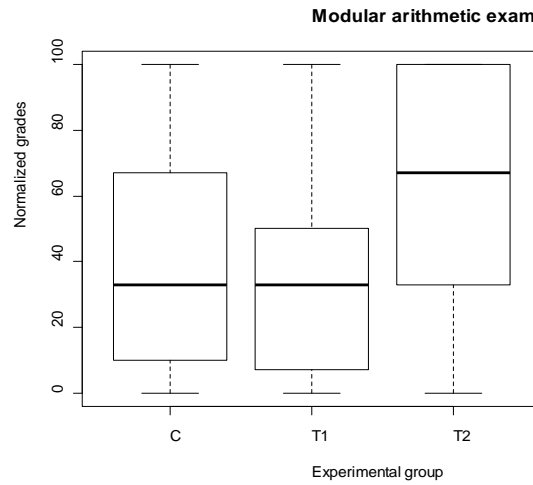


Figure 10. Box plot of the normalized grades obtained in each experimental groups for the modular arithmetic exam

To analyze if the treatments have produced any significant effect, the Kruskal-Wallis test is first performed. Results (see Table 1) indicate that there is no difference in the means for the grades of the real security case analysis ( $p$ -value  $> \alpha$ ) but there is a significant difference in the case of the modular arithmetic exam ( $p$ -value  $< \alpha$ ) for at least one pair of groups. Differences in variances cannot be inferred from data for any of the analyzed sets according to the results of the Fligner-Killeen test (see Table 2), as  $p$ -value  $> \alpha$  in both cases.

As the Kruskal-Wallis test indicates the existence of significant differences for the grades of the modular arithmetic exam, a Mann-Whitney test is performed for each pair of experimental groups to determine the specific cases in which the differences are significant. Results, summarized in Table 3, indicate that only treatment of group T2 (that is, the use of the trainer followed by the participation in the tournament) has a significant effect on student outcomes with respect to those obtained in both groups C and T1 ( $p$ -value  $< \beta$ ). However, results also show that treatment of group T1 (the use of the trainer) does not produce significant effects on student outcomes in comparison with the control treatment. This result partially contradicts the initial expectations.

Table 1. Results of the Kruskal-Wallis test on the analyzed grades for all experimental groups

Grades of the real security case analysis			Grades of the modular arithmetic exam		
Chi-squared	df	p-value	Chi-squared	df	p-value
0.2738	2	0.872	11.55	2	0.003104



Table 2. Results of the Fligner-Killeen test on the analyzed grades for all experimental groups

Grades of the real security case analysis			Grades of the modular arithmetic exam		
Med chi-squared	df	p-value	Med chi-squared	df	p-value
1.0815	2	0.5823	1.8326	2	0.4

Table 3. Results of the Mann-Whitney test performed on the grades for each pair of groups

Comparison	W	p-value
C vs T1	1114	0.4747
T1 vs T2	504	<b>0.001033<sup>(1)</sup></b>
C vs T2	731	<b>0.01154</b>

A jitter is applied to calculate an exact  $p$ -value.

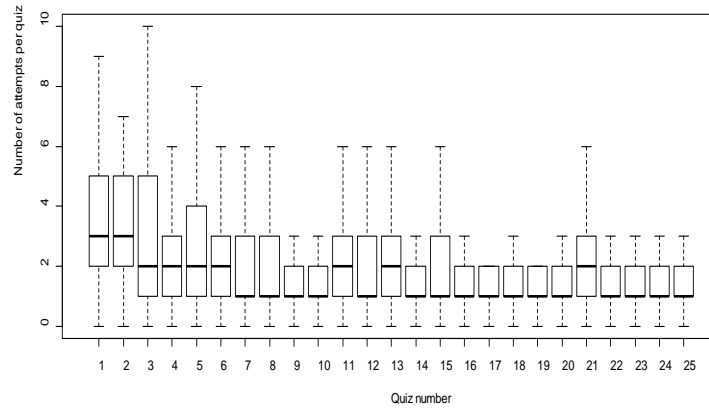


Figure 11. Box plot of the number of attempts for each quiz. Outliers have not been depicted.

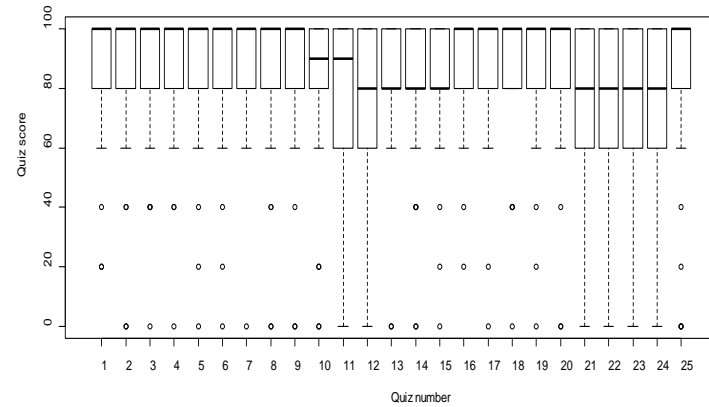


Figure 12. Box plot of the quiz scores for each quiz

On one hand, results are coherent with one of the initial expectations of this research. That is, doing more exercises may produce a stronger reinforcement of the targeted skills. The competition factor and the desire of winning any of the prizes engage students in the activity, as known by the researchers in conversations with students and shown by collected data (see Figure 11). Note that only five and four students were excluded from the study in groups T1 and T2, respectively, because of their low participation.

Although full knowledge of response feedback is not provided, students' knowledge of their own performance and confidence (after achieving a satisfactory level of correct answers in the quizzes) increases as they participate in the tournament. As shown in Figures 11 and 12, the number of attempts decreases as the tournament takes place, and the scores were rather high in all the quizzes of the tournament. The slightly lower scores in the quizzes of contests three and five are attributed to two reasons. The first reason is that the most difficult quiz bank was used in both cases,

because it considered the highest values of modulo  $n$  and, as a consequence, also of values  $a$  and  $b$ . The second purported reason is that, as the tournament contests took place, students' intrinsic motivation (the natural desire of winning the competition, the excitement created by the novelty of the experience, and the challenge posed to the students) decreased, and their participation in the tournament was mainly motivated by the effect that their score will have on their grades. It is presumable that longer tournaments (with contests taking place in more than five consecutive days) would have had a lower student participation in the end and a negative effect on student motivation.

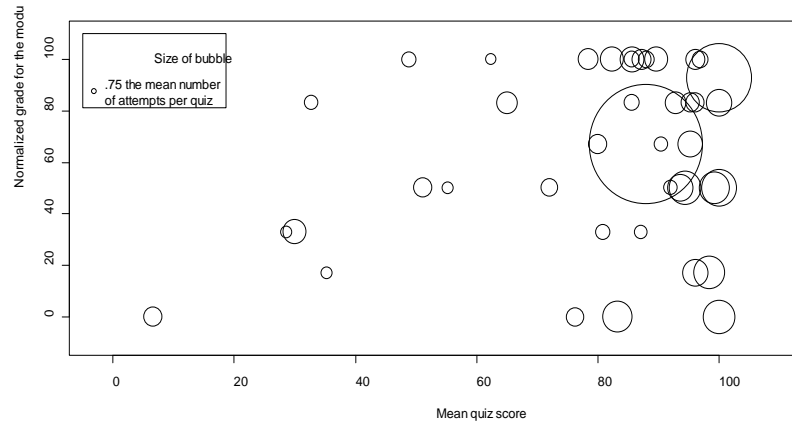


Figure 13. Bubble plot of the normalized grades for the modular arithmetic exam and the mean quiz score. The number of attempts per quiz determines the size of the bubble

The trainer was not intended to substitute homework sheets, but we expected that their use would produce slightly higher student outcomes because the students had the possibility of using a step-by-step trainer at their own pace, even out of class time. The result might be justified because higher education students, especially students of informatics engineering, are very used to computer-based technologies. Their use, however, does not increase students' motivation to learn compared to traditional lecture-based approaches. The trainer cannot compete with the rest of tasks assigned to students in other courses nor engage them significantly, in contrast to what appears to have happened with the tournament. The trainer, as it is designed, might be better than the lecture-based approach to make the students acquire the basics of the targeted skill (data collected in this experiment does not allow analyzing of this issue), but results of this experiment show that by itself, the trainer is not effective in reinforcing this learning.

Because results of the experiment indicate that the participation in the tournament affects the student outcomes, it is also interesting to analyze if there is any clear relation between the tournament score and the student's grade. As shown by Figure 13 and the small value of the Pearson's correlation coefficient ( $r = 0.2496$ ,  $p\text{-value} = 0.1108$ ), although the majority of students have a high score and most of them also a rather high grade, a high score does not imply a high grade. There are students that have a high score but a low grade.

On one hand, as it was known by the researchers in informal conversations with students (corridor talks held after lectures), some of these cases corresponded with students who choked when faced with the exam questions, and even they were surprised by their low performance in this situation. On the other hand, we noticed that few students have a low score and a high grade, reinforcing the previous conclusion of that participation in treatment 2 helps to reinforce the targeted skills and obtain higher grades.

Tables 4 and 5 summarize the results of the satisfaction survey. The number of students participating in the survey was 35 out of 41 in group T1 and 35 out of 42 in group T2. Data indicates that both groups of students were rather satisfied with the new instructional methods and that there are no significant differences between the perceived overall satisfaction level of both groups ( $p\text{-value} > \alpha$ ). However, it is remarkable that in both cases students' preference for the new instructional method with respect to the lecture-based approach presents the lowest values (3.54 and 3.22, respectively, for groups T1 and T2. Values are in bold in the table). This is because students are forced to learn actively instead of just attending the lectures and because they are not aware of the benefits of the tournament. We also noticed that, although not statistically significant (Kruskal-Wallis tests show  $p\text{-value} > \alpha$ ), fun and preference factors present the highest differences in the means, with the lower values belonging to group T2.

This can be an indication of that the competition, while having been softened, still has some slightly larger negative effect on students in comparison to non-competitive approaches. However, it could be also attributed to the fact that students who followed treatment 2 must actually work significantly more in comparison with those who followed treatment 1.

*Table 4. Results of the satisfaction survey for groups T1 and T2*

	T1		T2		Difference of means (% T2 – T1)
	Mean	SD	Mean	SD	
Effectiveness (1–5)	3.96	0.58	3.72	0.67	–6%
Easiness (1–5)	3.86	0.89	3.79	0.55	–1.8%
Motivation (1–5)	3.82	0.77	3.99	0.53	4.3%
Fun (1–5)	4.18	0.98	3.87	0.84	<b>–7.8%</b>
Preference (1–5)	<b>3.54</b>	1.26	<b>3.22</b>	0.81	<b>–8%</b>
Reuse (1–5)	3.86	1.01	3.79	0.76	1.8%
Student’s satisfaction (6–30)	23.214	3.645	22.38	2.739	–3.5%

*Table 5. Results of the Kruskal-Wallis and Fligner-Killeen tests on the students’ satisfaction for both treatment groups T1 and T2*

Kruskal-Wallis test			Fligner-Killeen test		
Chi-squared	df	p-value	Med chi-squared	df	p-value
0.8445	1	0.3581	0.1877	1	0.6648

Results of this study are very similar to those obtained by Regueras et al. (2009), which is a logical result as the design of the competition of both experiments share two main characteristics, detailed next. First, the competition is a multiple-prize soft competition, i.e., it is not a pure interpersonal competition. In QUEST, the system analyzed in (Regueras et al., 2009), this is instrumented by the following:

- making the students to compete in pairs
- adjusting question scores depending on the time elapsed from its publication
- allowing students to propose challenges
- awarding prizes in a proportional way depending on the students’ position in the rank
- allowing access to the system through Internet (without any other restriction)

By contrast, in the system proposed and analyzed here, the softness of the competition is instrumented by (i) allowing multiple attempts for the same quiz, (ii) adjusting quiz scores depending on the number of attempts, (iii) making public the quiz scores (and not the weighted quiz scores), (iv) awarding prizes in a proportional way depending on the students’ position in the rank and a fixed set of prize categories with a limited quota, (v) rewarding the majority of participating students (a very small successful participation allows winning something), and (vi) allowing participation in the tournament through Internet. Second, the competition consists of several steps combined with public ranks, so students have the opportunity to react to results published in the competition ranks. Competition games analyzed in Cameron & Dwyer (2005) and Ebner & Holzinger (2007) do not present one or both of these features (see Figure 1).

Differences in the outcomes between treatment and control groups in this work and in Regueras et al. (2009) can be attributed to the fact that students actually do additional exercises (in our case, motivated externally by the explicit reward and internally by the competition), which sounds like a reasonable hypothesis. Note that in our case, we presume that students of control and treatment 1 groups have not put too much effort into the optional reinforcement homework drills. However, attributing results to doing additional exercises cannot be generalized without considering other factors, as it is refuted by Ebner and Holzinger (2007). In their experiment, non-playing students did not perform additional exercises either, but results indicate that no significant differences exist between playing and non-playing students.

## Conclusions

In Higher Education Engineering studies, optional drill and practice activities have the risk of students' not fulfilling the activity or not putting enough effort on it, thus losing effectiveness despite being crucial for students learning to think mathematically. Then, there exists the need of engaging students in performing this type of activity in the context of reinforcing routine mathematical skills.

Although competitive approaches are not a trend in current educational practices, there are several arguments that support the use of competitive approaches in higher education for engaging students in certain learning activities. However, negative effects of pure interpersonal competition need to be lessened. Recent works in this direction have obtained mixed results both in pre-higher and higher education, showing the need to further contribute in this direction.

In this work, results of a quasi-experiment conducted to evaluate the effects on students' outcomes and satisfaction of a competitive web-based drill game are presented. As well, a novel, multiple-prize soft competition design that considers several stages combined with public ranks is proposed, with the goal of minimizing the disadvantages of pure interpersonal competitions. Results indicate that the tournament is highly effective in increasing student outcomes and that student satisfaction after the experience is rather high. A trainer system consisting of a set of web-based quizzes with elaborative feedback complements the tournament. Use of the trainer alone has also been evaluated. Results show that the trainer system does not present significant differences for reinforcing the targeted skills compared to a lecture-based approach. The level of satisfaction of students who used only the trainer is similar to the level of satisfaction of students who also participated in the tournament.

In light of previous results and taking into account results of related work, we can conclude that multiple-prize soft competitions involving several steps combined with public ranks are a key factor for obtaining significant results. Nonetheless, the following limitations exist:

- More experiments with different soft competition designs and different contexts are needed to confirm the conclusion.
- The existence of an explicit reward may be influencing the results, because experiments that did not consider the reward did not obtain clear, significant results. Note, however, that some works refute this hypothesis in other contexts (Blanes i Vidal & Nossol, 2011). Experiments that analyze the effects of this variable are also needed.
- The proposed prize distribution structure requires a high workload to calculate the rewards. It has been even necessary to count with supportive software. It should be worth investigating which are the more efficient competition designs regarding this aspect.

Future work will try to address these questions and also to extend the methodology to other modules of the course that share the same characteristics. Furthermore, students of future courses will be informed about the benefits of participating in the tournament to gradually decrease the current effect of the tournament score on the student grade (i.e., substitute an external motivation by an internal one).

## Acknowledgements

This work was supported by the University Carlos III of Madrid under the 2009/2010 Convocatoria de Apoyo a Experiencia de Innovación e Internacionalización Docente. The authors would also like to acknowledge the assistance of Miguel Angel Valentín, José Miguel Ballesteros, Adrián Tabara, Guillermo Suarez-Tangil, and Jorge Blasco.

## References

- Blanes i Vidal, J., & Nossol, M., (2011). Tournaments without prizes: Evidence from personnel records. *Management Science*, 57(10), 1721-1736. doi: 10.1287/mnsc.1110.1383
- Cameron, B., & Dwyer, F., (2005). The effect of online gaming, cognition and feedback type in facilitating delayed achievement of different learning objectives. *Journal of Interactive Learning Research*, 16(3), 243-258.

- Cason, T. N., Masters, W. A., & Sheremeta, R. M., (2010). Entry into winner-take-all and proportional prize-contests: An experimental study. *Journal of Public Economics*, 94(9), 604–611.
- Conference of European Ministers Responsible for Higher Education (2009, April ). The Bologna process 2020: The European higher education area in the new decade. *Communiqué of the Conference of European Ministers Responsible for Higher Education, Leuven and Louvain-la-Neuve, 28–29 April 2009*. Retrieved from [http://europa.eu/rapid/press-release\\_IP-09-675\\_en.htm](http://europa.eu/rapid/press-release_IP-09-675_en.htm)
- CS2008 Interim Review Task Force of the ACM and the IEEE Computer Society (2008). *Computer Science Curriculum 2008: An interim revision of CS 2001*. Retrieved July 24, 2012 from: <http://www.acm.org/education/curricula/ComputerScience2008.pdf>
- Ebner, M., & Holzinger, A., (2007). Successful implementation of user-centered game-based learning in higher education: An example from civil engineering. *Computers & Education*, 49(3), 873–890.
- Johnson, R.G., Johnson, D.W., & Stanne, M.B., (1985). Effects of cooperative, competitive, and individualistic structures on computer-assisted instruction. *Journal of Educational Psychology*, 77(6), 668–677.
- Ke, F., & Grabowski, B., (2007). Gameplaying for maths learning: Cooperative or not? *British Journal of Educational Technology*, 38(2), 249–259.
- Kebritchi, M., Hirumi, A., & Bai, H., (2010). The effects of modern mathematics computer games on mathematics achievement and class motivation. *Computers & Education*, 55(2), 427–443.
- Lam, S., Yim, P., Law, J.S.F., & Cheung R.W.Y., (2004). The effects of competition on achievement motivation in Chinese classrooms. *British Journal of Educational Psychology*, 74(2), 281–296.
- Moldovanu, B., & Sela, A., (2001). The optimal allocation of prizes in contests. *The American Economic Review*, 91(3), 542–558.
- Regueras, L. M., Verdú, E., Muñoz, M. F., Pérez, M. A., de Castro, J. P., & Verdú, M. J., (2009). Effects of competitive e-learning tools on higher education students: A case study. *IEEE Transactions on Education*, 52(2), 279–285.
- Rzoska, K.M., & Ward, C., (1991). The effects of cooperative and competitive learning methods on the mathematics achievement, attitudes toward school, self-concepts and friendship choices of Maori, Pakeha and Samoan Children. *New Zealand Journal of Psychology*, 20(1), 17–24.
- Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition and sense making in mathematics. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning: A project of the National Council of Teaching of Mathematics* (pp. 334–370). New York, NY: Macmillan.
- Smith, K.A., Sheppard, S.D., Johnson, D.W., & Johnson, R.T., (2005). Pedagogies of engagement: Classroom-based practices. *Journal of Engineering Education*, 94(1), 87–101.

## Assessment of Animated Self-Directed Learning Activities Modules for Children's Number Sense Development

Der-Ching Yang<sup>1\*</sup> and Mao-Neng Li<sup>2</sup>

<sup>1</sup>Graduate Institute of Mathematics and Science Education, National Chiayi University, Taiwan, R.O.C. //

<sup>2</sup>Department of Education, National Chiayi University, Taiwan, R.O.C. // dcyang@mail.ncyu.edu.tw // fredli@mail.ncyu.edu.tw

\*Corresponding author

(Submitted April 02, 2012; Revised November 16, 2012; Accepted December 25, 2012)

### ABSTRACT

The purpose of this study was to examine the relative effectiveness of two different learning modes; namely, a computer animation self-directed learning approach and a paper version of the self-directed learning approach, to 5<sup>th</sup>-graders' number sense development. Two 5th-grade classes, 30 students each, were selected from a public elementary school in southern Taiwan and randomly assigned to the computer animation experimental group [CAEG] and the paper-version group [PAG]. The major finding indicated that students in the CAEG group had better performance on number sense and showed more frequent uses of number sense than students in the PAG group. Implications for the uses of computer animation self-directed learning activities are discussed.

### Keywords

Computer animation activities, Learning modes, Number sense, Self-directed learning approach

### Introduction

Number sense has been considered as one of the most important topics in mathematics education (Dunphy, 2007; Jordan, Glutting, & Ramineni, 2010; National Council of Teachers of Mathematics, [NCTM], 2000; Verschaffel, Greer, & De Corte, 2007; Yang & Tsai, 2010). However, several number sense studies showed that not only middle and elementary school children (Markovits & Sowder, 1994; McIntosh, B. Reys, R. Reys, Bana, & Farrel, 1997; Menon, 2004a; Reys & Yang, 1998; Yang & Li, 2008) but also preservice teachers performed poorly in solving problems related to number sense (Menon, 2004b; Yang, Reys, & Reys, 2009). Mathematics educators have found that children's inflexibility in handling number sense problems may be related to the inappropriate design of mathematics textbooks and the rule-based instruction of school mathematics (Markovits & Sowder, 1994; Menon, 2004a; Reys & Yang, 1998; Yang & Li, 2008). For example, previous studies showed that traditional instruction, which focuses primarily on standard written computation rather than conceptual understanding, may result in children's difficulties in using number sense flexibly and efficiently (Menon, 2004a; Yang & Li, 2008). Helping children develop number sense has been highlighted in school mathematics by many reports and studies (Bobis, 2004; Dunphy, 2007; McIntosh et al., 1997; NCTM, 2000). Seeking an effective learning material that would facilitate children's development of number sense is an urgent issue for current mathematics education.

Previous studies and reports have strongly suggested that technology should be integrated into mathematics teaching and learning, since technology is a promising means for providing conceptual understanding (Dick, 2007; Inamdar & Kulkarni, 2007; NCTM, 2000; Ruthven, 2007; Vulis & Small, 2007; Yang & Tsai, 2010; Zbiek, Heid, Blume, & Dick, 2007). For example, NCTM (2000) highlights that the integration of technology into mathematics classroom settings can enhance children's mathematics learning and promote conceptual understanding. In addition, Olive and Lobato (2008) state that "providing technological tools that enable children to enact their own mathematical operations can be a very powerful aid to children's learning" (p. 43). Technology can provide dynamic visualization, immediate feedback, and interactivity that can help children learn mathematics effectively (Gegner, Mackay, & Mayer, 2009; Suh & Moyer-Packenham, 2007). Therefore, it is reasonable to believe that technology should play a key role in helping children learn mathematics.

Accordingly, the purpose of this study was to examine the effectiveness of computer animation self-directed learning number sense-based activities compared with a corresponding paper-version approach. The specific research questions are as follows:

- Are there any significant differences in students' performance on number sense across the two self-directed learning approaches after the experiment?
- Are there any differences in methods used by students to solve number sense problems across the two self-directed learning approaches after the experiment?

## Background

### Number sense framework

What is number sense? Number sense refers to an individual's general understanding of numbers, operations, and the relationship between number and operation, and the ability to handle daily-life situations that include numbers (Markovits & Sowder, 1994; Reys & Yang, 1998; Yang & Tsai, 2010). During the past two decades, due to the importance of number sense, several research studies and reports have been produced (Markovits & Sowder, 1994; NCTM, 2000; Reys & Yang, 1998; Verschaffel et al., 2007; Yang & Li, 2008; Yang & Tsai, 2010). Even though these studies and reports defined number sense from different perspectives, they have included the following major common components: (a) Understanding the basic meanings of numbers; (b) recognizing number size; (c) Being able to use multiple representations; (d) Recognizing the relative effect of operation on numbers; and (e) Being able to judge the reasonableness of a computational result via estimation or mental computation (Markovits & Sowder, 1994; NCTM, 2000; Reys & Yang, 1998; Verschaffel et al., 2007; Yang & Tsai, 2010). These common components of number sense were used in this study.

### The significance of number sense in elementary school mathematics education

Why is number sense development for elementary school students so important? There are three reasons for its significance. First, number sense is a way of thinking that promotes sense-making and directs problem solving in a flexible and efficient way (Dunphy, 2007; Yang & Li, 2008). For example, when children were asked to find the answer to  $25 \times 96 \div (16 \times 5)$ , children tended to use paper-and-pencil to compute the answer:  $25 \times 96 = 2400$ ,  $16 \times 5 = 80$ , and then  $2400 \div 80 = 30$ . It is often difficult for children to discover the relationships among these numbers, for example, firstly, we can mentally compute  $25 \div 5 = 5$  and  $96 \div 16 = 6$  and then  $5 \times 6 = 30$ , which is more flexible and efficient way to find the answer. Hence, the answer could be found easily:  $5 \times 6 = 30$ . Second, number sense plays a key role in mathematics achievement for elementary school children. For example, Jordan, Glutting, and Ramineni (2010) showed that number sense could make an important contribution to mathematics achievement. Moreover, the study of Yang & Li (2008) revealed that number sense is significantly correlated with fifth graders' mathematics achievement. Third, several studies have suggested that overemphasis on written computation may limit children's mathematical thinking and understanding, whereas number sense can induce a process of meaningful learning (Yang & Li, 2008). Therefore, it promotes conceptual understanding and flexible ways of problem solving. For example, when asking children to decide which of  $2/5$  or  $4/7$  is larger, students usually use paper- and-pencil (first finding the common denominator, calculating the numerators, and finally comparing the numerators), which is often a meaningless process. A child with good number sense should quickly discover that  $2/5 < 1/2$ , and  $4/7 > 1/2$ . Therefore, the answer is  $2/5 < 4/7$ .

### Computer technology and mathematics learning

The progress of computer technology has changed the learning style of human beings (NCTM, 2000; Ministry of Education in Singapore, 2009), including playing an important role in mathematics teaching, learning, and evaluation (Ruthven, 2007; Zbiek et al., 2007). Several studies have indicated that children's learning outcomes can be improved through the mode of virtual manipulative computer learning (Moyer, Niezgoda, & Stanley, 2005; Reimer, & Moyer, 2005; Suh & Moyer-Packenham, 2007). Mathematics virtual manipulation is defined as an interactive (acting and responding to each other), web-based, virtual representation of a dynamic object that presents opportunities for children to construct mathematical knowledge (Moyer, et al. 2005; Suh & Moyer-Packenham, 2007). Hence, children may have greater opportunity to learn when the manipulation occurs on computers, and that may advance their mathematical understanding (Chan, Tsai, & Huang, 2006; Isiksal & Askar, 2005; Yang & Tsai, 2010).

In recent years, many studies have shown that technology integrated into mathematics teaching and learning can efficiently and effectively help children develop better conceptual understanding of mathematics (Bennison & Goos, 2010; Dick, 2007; Ruthven, 2007; Suh & Moyer-Packenham, 2007). For example, Suh and Moyer-Packenham (2007) indicate that a virtual environment can enhance students' ability to connect visual and symbolic representations, support and direct the process of a mathematics algorithm step by step, and provide rapid feedback as well as opportunities for self-examination. Zbiek et al., (2007) indicate that technology-integrated mathematics

learning and instruction can promote students' representational fluency. Moreover, previous studies also suggested that multiple representations can help children develop better understanding of mathematics (Goldin, 2002; Tripathi, 2008). Hence, multiple representations with the applications of modern technology through dynamic visualization appear promising in children's mathematics learning.

### **Animation and number sense related studies**

Due to the importance of number sense, several studies have demonstrated that children's number sense can be promoted by integrating well-designed number sense activities into mathematics classrooms or by including realistic activities focusing on number sense (Griffin, 2004; Yang & Wu, 2010). Since technology has been regarded as an important and effective learning tool for helping children develop better understanding of mathematics (Bennison & Goos, 2010; Dick, 2007; Reimer & Moyer, 2005; Ruthven, 2007; Suh & Moyer-Packenhams, 2007), technology should be a relevant tool for promoting children's number sense. However, few studies via technology have focused on helping children develop number sense (Hudson, Kadan, Lavin, & Vasquez, 2010; Su, Marinas, & Furner, 2010; Yang & Tsai, 2010). For example, the study of Su et al. (2010) showed that web-based tools could be used to reinforce primary school children's number sense. In addition, the study of Yang and Tsai (2010) applied the technology-based environment to help teacher's teaching and student's learning in number sense. Their findings indicated that "integrating technology into number sense teaching and learning not only could promote students' number sense, but also had a positive effect on student's attitudes towards the improvement of number sense" (p. 112).

In addition, several studies suggested that students may learn more from multimedia-based learning environment due to multimedia messages including words and visuals than what traditional teaching modes can do (Gegner, Mackay, & Mayer, 2009; Mayer, 2003). Even though, Clark & Mayer (2008) state that "to illustrate how things work, a series of still visuals is likely to lead to as good or better understanding of the process than animation" (p. 8), unless carefully designed toward the fostering of conceptual understanding (Clark & Mayer, 2008). They provided several evidence-based studies to support their suggestions, but these experiments did not focus on mathematics related topics (Clark & Mayer, 2008; R. Mayer, Hegarty, S. Mayer, & Campbell, 2005). Furthermore, several previous studies (Forster, 2006; Philpot, Hall, Hubing, Flori, Oglesby, & Vikas, 2003; Stoffa, 2004; Taylor, Pountney, & Malabar, 2007) commented that animation might be an aid in mathematics teaching and learning and helping students understand mathematics. For example, the study of Forster (2006) indicated that it is essential to get technology integrated into the assessment, examinations, and classroom mathematics practices if we want to optimize students' mathematics learning. The findings of Taylor et al.'s (2007) experiment showed that animated learning materials could act as an aid in helping mathematics teaching and learning for students at undergraduate level. In addition, the results of these studies showed that animated learning materials including dynamic figures, words, and visuals could help children develop mathematical concepts. This indicates that the well-designed animation activities can be a more effective learning tool than static visualization.

Previous literature reviews have shown that animation can play an important role in students' mathematics learning. For example, Taylor et al. (2007) studied on undergraduate students' mathematics learning. Forster (2006) examined high school students' mathematics learning via animation activities. However, little research has been conducted regarding the use of animation to improve number sense at the primary school level. Consequently, this study was to investigate the effectiveness of animation to support number sense self-directed learning for fifth-graders, compared to static paper-version group and traditional textbook learning group. These animated activities emphasized understanding and potential applications of number sense in order to help children develop number sense in a dynamic and effective manner.

## **Method**

### **Research design**

A quasi-experimental pretest-posttest design was used to investigate the effects of different self-directed learning formats on children's development of number sense. Specifically, two intact classes were randomly assigned to a computer animation experimental group (CAEG) and a paper-version group (PAG). They learned number sense by themselves via different presentations of learning materials in a different form of organization without a teacher's



instruction. Teachers acted merely as a classroom manager and a material provider, ensuring that the student's self-directed learning opportunities ran smoothly. Each treatment group followed a different guided self-directed learning approach and was given the same amount of self-directed learning time. Each treatment condition was implemented during seven class periods (30 minutes each) over consecutive 4 weeks. The pretest was given in the 1<sup>st</sup> week. During 7 class periods (30 minutes per period) in the next 2 weeks, self-instruction focused on each of the learning formats using different structural modules of instructional materials. The posttest was administered in the last week.

## Participant

Sixty students in two fifth-grade classes were selected from an elementary school in Taiwan. These students' average age was about 10.8 years old. The two classes were randomly assigned to one of the self-directed learning groups (CAEG and PAG). The CAEG group included 30 students (16 boys and 14 girls) and the PAG group included 30 students (16 boys and 14 girls).

## Treatments

Two independent variables, self-directed learning materials-presentation formats and self-directed learning approaches were manipulated across each treatment condition. There were two formats for presenting self-directed learning materials: the animation version (for CAEG) and the paper version (for PAG).

To help participants develop number sense, seven self-directed learning units were designed for different types of self-directed learning. One activity from each unit was provided during each of the seven class periods. For the CAEG group, seven computer animated self-directed learning activities based on the lessons in the textbook used at the elementary school, were designed by the researchers and were number sense oriented (Griffin, 2004; Yang & Wu, 2010). These seven animated self-directed learning activities focused on conceptual or relational understanding. During the processes of animation activities, one example was presented through dynamic visualization, audio, interaction, and questions to lead students in learning number sense related concepts. Immediate feedbacks on students' selections were presented on the screen. If student answers were incorrect, the screen showed "Your answer is incorrect, please try it again!" After the example, students were asked to take a quiz which includes several questions. The testing results were displayed to let the students know their performance. Take fraction comparison for example, one of the examples used for CAEG group is described as follows:

A3. Speed eating activity for CAEG: Mary ate  $\frac{8}{9}$  of a cake and John ate  $\frac{15}{16}$  of the same size of the cake. Who ate more cake?

For the CAEG group, a series of computer animated learning activities for problem A3 is shown and explained clearly in Figure 1.



Note. <sup>a</sup> indicates visual display of the item in Chinese characters (English version: Which fraction is the best to represent the remained part of cake not eaten by Jane?)

<sup>b</sup> indicates an item options.

<sup>c</sup> indicates visual display of the item in a picture.

The following example pictures are presented in a similar way.



Which fraction is the best to represent the remained part of cake not eaten by Tom?



My smart kids, can you tell me whose remained piece of cake is larger, Jane or Tom?



$1/9$  implies that Jane takes one piece of cake out of a cake which is equally divided into 9 pieces.  
 $1/16$  implies that Tom takes one piece of cake out of a cake which is equally divided into 16 pieces.  
 Therefore,  $1/9 > 1/16$ .

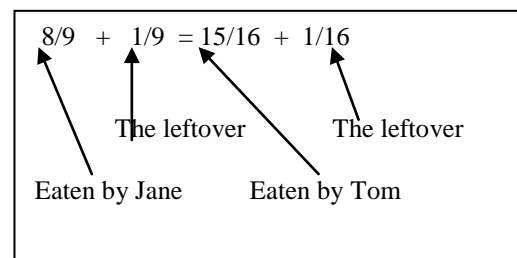


Figure 1. The learning activity of problem A3 for the CAEG Group

Figure 1 presents the question posed on the screen. A piece of a cake was left and Jane ate the other pieces as shown in the first picture. Which fraction is best to represent the left part? A piece of a cake was left and Tom ate the other pieces as shown in the second picture. Which fraction is best to represent the leftover part?


The seven paper-version activities for the PAG group were transferred from the seven computer animated activities created for the CAEG group and they were also number sense oriented. As to the seven self-directed learning units for the students in the PAG group, they were literally just a printed version of the computer-based materials, including the same visual display of the item in Chinese words, the item options, and the same visual display of the item in a picture, with no moving animations. The contents for the activities in both groups were the same except the activities used in the PAG were individually exposed to self-directed learning activities printed on the paper, without any animation. Contrast, one of the corresponding items used for the CAEG group is exemplified as follows:

For the PAG group, the corresponding paper-version of the computer animated learning activities for problem A3 is shown in Figure 2. Those questions used in CAEG group were also presented in Figure 2.

a 裁判二：「如右圖，小英剩下的是一個蛋糕的幾分之幾呢？」

b (1)  $\frac{1}{9}$  (2)  $\frac{8}{9}$  (3)  $\frac{9}{8}$  (4) 8

我的答案：\_\_\_\_\_

c 

Note. <sup>a</sup> indicates visual display of the item in Chinese characters (English version: Which fraction is the best to represent the remained part of cake not eaten by Jane?)

<sup>b</sup> indicates the item options.

<sup>c</sup> indicates the visual display of the item in a picture.

The following example pictures are presented in a similar way

裁判二：「小華剩下的是一個蛋糕的幾分之幾呢？」

(1)  $\frac{1}{16}$  (2)  $\frac{15}{16}$  (3)  $\frac{16}{15}$  (4) 15

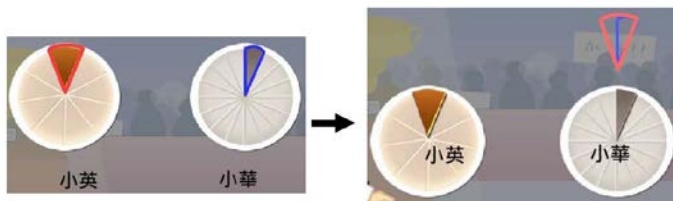
我的答案：\_\_\_\_\_



Which fraction is the best to represent the remained part of cake not eaten by Tom?

My answer is \_\_\_\_\_.

裁判二：「你有沒有發現了什麼呢？根據移走未吃完的蛋糕，我們發現，小英未吃完的蛋糕是  $\frac{1}{9}$ ，小華未吃完的蛋糕是  $\frac{1}{16}$ 。聰明的小朋友，你可以從下圖看得出來哪一個人剩下來的蛋糕比較大嗎？」



(1) 小英未吃完的蛋糕較大 ( $\frac{1}{9} > \frac{1}{16}$ )

(2) 小華未吃完的蛋糕較大 ( $\frac{1}{9} < \frac{1}{16}$ )

(3) 小英未吃完的蛋糕與小華未吃完的蛋糕一樣大 ( $\frac{1}{9} = \frac{1}{16}$ )

我的答案：\_\_\_\_\_

Please look at the uneaten part of cake!

My smart kids, can you tell me whose remained piece of cake is larger, Jane or Tom?

- (1) Jane
- (2) Tom
- (3) Identical

裁判二：「哦～因為 $\frac{1}{9} > \frac{1}{16}$ ，表示小英未吃完的部份 $\frac{1}{9}$

比小華未吃完的部份 $\frac{1}{16}$ 多，所以你想到了嗎？

因為小英剩下的部分比小華多，

所以小華吃掉的比小英吃掉的多！

啊～原來我們只要比較剩下的部份就可以很快的找出答案了耶！

$$\left(\frac{8}{9} + \frac{1}{9}\right) = 1 = \left(\frac{15}{16} + \frac{1}{16}\right)$$

那如果以後出現數字比較複雜或是沒有圖形表示的時候，  
我們就可以用這個方法很快的找出答案耶！」

$1/9$  implies that Jane takes one piece of cake out of a cake which is equally divided into 9 pieces.  
 $1/16$  implies that Tom takes one piece of cake out of a cake which is equally divided into 16 pieces.  
Therefore,  $1/9 > 1/16$ .

$$\frac{8}{9} + \frac{1}{9} = \frac{15}{16} + \frac{1}{16}$$

Figure 2. The learning activity of problem A3 for the PAG. group

The contents of the self-directed learning activities across two different self-directed learning groups were the same. The major difference between the two groups was the self-directed learning formats of the activities used for each group. For the CAEG group, they learned number sense by themselves through seven animated number sense-based activities, which were designed by the researchers. Each animated self-directed learning activity consisted of a series of animated pictures. Each student individually completed these animated activities inside a computer classroom. The main themes of the seven activities are described in Appendix 1.

## Instruments

The pretest and the posttest involved the same test; namely, a web-based, two-tiered test for fifth-grader's number sense, designed by the Yang (2010). This test includes five number sense components as defined in the number sense framework with eight items for each component. The value for Cronbach's  $\alpha$  was .80. The difficulty indices for the test items were set between .28 and .68, and the discrimination indices were between .25 and .77. Additionally, the construct reliability index derived from SEM analysis was .801, reflecting that the instrument's internal and construct reliability were satisfactory.

The test's examined validity covered the content validity, specialist validity, and construct validity. To ensure the designed items were representative and not beyond the curriculum scope usually taught to 5<sup>th</sup>-graders in Taiwan, three elementary school teachers and two mathematics educators were invited to review the items. All agreed that the 40 test items were representative and appropriate in terms of item content for 5<sup>th</sup>-graders involved in this study. An exemplified item is presented in Appendix 2.

## Procedure

This study was conducted in three stages:

- During the pretest, all of the students were asked to individually complete the 50-minute, computerized pretest in a computer classroom using the web-based testing platform.
- During the learning treatment stage, the students in the CAEG group were asked to individually complete the computer animated units in the computer classroom. The students in the PAG were asked to individually read the seven units of the paper-version that were adapted from the computer-animated units and thus included the same content as the units studied by the students in the CAEG.
- The posttest was conducted 1 week after the learning treatment ended. During the posttest, all of the students were asked to complete individually the 50-minute, computerized posttest in a computer classroom using the same web-based testing platform.

## Scoring and data analysis

The web-based, two-tier test for number sense was composed of 40 multiple-choice items. For each test item, the students were required to provide their choice for the correct answer and a reason for their answer selection. This scoring design improves score validity because it identifies genuine understanding of the problem. Due to the unique features of this test, the scoring scheme for the pretest and the posttest is detailed as follows (see Table 1):

Table 1. Scoring Scheme for the Number Sense Test

Answer	Correct 4 points	4 points	4 points	4 points	Wrong 0 point
Reason	NS-Based 4 points	Rule-Based 2 points	Misconceptions 1 point	Guess 0 points	
Total	8 points	6 points	5 points	4 point	0 points

We believe that the correct reason (reflecting number sense application) is of equal importance as that of the correct answer, because appropriate use of number sense is a good indicator of effective mathematical thinking. Therefore, students receive the same number of points for each: 4 points for the appropriate application of number sense and 4 points for the correct answer.

## Results

### Learning effect on different learning modes

Table 2 reports the descriptive statistics and the results of the repeated  $t$  test between the pretest and the posttest for the two groups. Results indicate that there is a statistically significant difference between the pretest and the posttest ( $t$  value =  $-6.401$ ,  $p < .01$ ) for the computer-animated experimental group (CAEG). The mean gain from pretest to posttest for the CAEG group was 33.46, indicating a positive effect on students' number sense performance. The data also show that the mean scores on the posttest for the PAG increased 5.84 points. The repeated  $t$ -test results indicate that there was no statistically significant difference between the pretest and posttest means for the PAG group ( $t = -1.717$ ,  $p > .05$ ). These results indicate that number sense learning via integrated technology is more beneficial to the students in the CAEG group than students learning number sense without the aid of technology in the PAG group.

Table 2. Descriptive Statistics and Repeated  $t$ -tests

Group	Pretest Mean (SD)	Posttest Mean (SD)	$t$	$p$
CAEG ( $n=30$ )	137.37 (53.74)	170.83 (54.49)	-6.401	.000**
PAG ( $n=30$ )	131.63 (47.16)	137.47 (53.30)	-1.717	.097

\*\* $p < .01$

In order to reduce the threats of initial pre-existing difference between groups and self-growth, an ANCOVA for the two groups was conducted with the pretest as a covariate. Before the ANCOVA was done, the homogeneity test of equal slopes between groups was examined. Because  $p > .05$ , the result  $F_{(1, 58)} = 3.629$  ( $p = .062$ ) indicates that the assumption of equal slopes is tenable, thus permitting the use of ANCOVA. Table 3 reports the results of the ANCOVA analysis.

Table 3. Summary of ANCOVA for Two Groups

Source	Type III Sum of Squares	$df$	Mean Square	$F$	$p$	Partial $\eta^2$
Pre	134993.821	1	134993.821	229.624	.000**	.801
Group	11633.641	1	11633.641	19.789	.000**	.258
Error	33509.812	57	587.891			

Note.  $R^2 = 0.819$  (adjusted  $R^2 = 0.813$ ).

\*\* $p < .01$

After adjusting the pre-existing difference between groups, the result of the ANCOVA is an  $F$ -value equal to 19.789 ( $p = .000$ ,  $\eta^2 = .258$ ). This result indicates that there is a large effect ( $\eta^2 > .14$ , based on Cohen's criterion) and also a statistically significant difference ( $\alpha = .05$ ) between the CAEG and PAG groups after the experiment. In other words, after the individual self-directed learning (using animation vs. paper version), the students in the CAEG group using computer-animated activities performed better on number-sense items than students in the paper-version group. In fact, participation in the computer-animated experimental group greatly enhanced the development of students' number sense ability. In addition, the  $R^2$  value (.819) indicates that the individual learning accounted for 81.9% of the total variation in the learning effect for the computer-animated experimental group. This implies that the technology integrated into the individualized number-sense learning had a significant impact on number sense performance, surpassing the self-directed learning through paper version.

### Problem-solving methods used within different learning modes

Table 5 reports the percentages of different methods of problem solving selected by students from different groups before and after the treatment. Data suggested that uses of the number-sense-based method increased much more for students in the CAEG (35.4% - 25.3% = 10.1%) than for students in the PAG (25.1% - 23.9% = 1.2%) group. This suggests that computer-animated learning is more effective than a paper-version learning approach in the development of number sense. At the same time, data also showed that the students in the CAEG group with incorrect answers decreased from 50% to 38.6% after the experiment. However, there is just a little decrease for students with incorrect answers from 51.7% to 49.2% in the PAG group after the experiment.

Table 5. Percentage of Different Methods Used to Solve Test Items Before and After the Experiment

Methods	Groups	CAEG		PAG	
		Pre-test	Post-test	Pre-test	Post-test
Correct	NS-based	25.3%	35.4%	23.9%	25.1%
	Rule-based	14.2%	15.1%	14.1%	15.7%
	Misconception	10.2%	10.8%	9.2%	8.6%
	Guessing	0.3%	0.1%	1.1%	1.4%
Incorrect		50.0%	38.6%	51.7%	49.2%

### Discussion

This study provided evidence of more effective self-directed learning of number sense through computer-animated activities with fifth graders in a public elementary school in southern Taiwan. The most important finding was that students in the CAEG group that received self-directed learning computer-animated number sense activities performed better on a number sense assessment than students in the PAG group who read a paper-version activities based on the same content as the computer-animated activities. This implies that children can develop number sense through self-directed learning in animation-based environments. This finding supports the statement of the NCTM (2000) that "technology can help students learn mathematics" (p. 25) and "students' engagement with, and ownership of, abstract mathematical ideas can be fostered through technology" (p. 25). It also supports the findings of several earlier studies that technological tools can help to foster children's mathematics learning (Chan et al., 2006; Dick, 2007; Inamdar & Kulkarni, 2007; Zbiek et al., 2007). Moreover, these findings also suggest that students in the CAEG group were more effective in their uses of number sense-based skills in solving problems than the students in the PAG group. Interestingly, we found that the students in the CAEG group who individually manipulate the animation activities applied much more uses of number sense than the students in the PAG group after the experiment. Rather, they highly decreased the incorrect responses than the students in the PAG group after the experiment. This finding strongly supports the results of an earlier study of Su et al. (2010) that web-based tools could be used to reinforce primary school children's number sense. It also supports the finding of Yang and Tsai (2010) that student's number sense could be improved through the technology-based environment.

The students in the CAEG group performed better on number sense than students in the PAG group because students in the CAEG group have developed better conceptual understanding on number and operations under animated-based learning environment that included clearer words and pictures than the paper-version group mode. The key



factors of CAEG group have better performance and use of number sense due to these animation activities bring dynamic visualizations, interactivity, and immediate feedback. This finding is consistent with the earlier studies (Gegner et al., 2009; Mayer, 2003). In fact, this study also proved that animated and number sense-based self-directed learning materials may be more effective than static visuals to promote children's number sense. These animated learning materials included dynamic visuals and words that made an appeal to student's attentions when they were manipulating these activities. This finding also supports previous studies that animation activities can be an aid in mathematics learning and helping students develop better understanding on mathematical concepts (Forster, 2006; Philpot, Hall, Hubing, Flori, Oglesby, & Vikas, 2003; Stoffa, 2004; Taylor, Pountney, & Malabar, 2007). Furthermore, in reviewing the number sense-based animation activities used in this study, they are consistent with the Mayer's principles when learning is better (Clark & Mayer, 2008). For example, these activities include relevant visuals and words rather than words only (Multimedial principle); texts in the activities have close connection with the visuals (Contiguity principle); visuals used in the activities are clearly explained by audio rather than by words (Modality principle); learners can mentally interact and get feedback via on screen learning agents (Personalization principle); key concepts of number sense are introduced first and then the exercises are presented (Pretraining principle); and contents are presented in a series of dynamic visuals which the learners can control the rates of access (Segmenting principle) (Clark & Mayer, 2008). Therefore, it is reasonable to believe that the finding of this study is different from the statement of Clark & Mayer (2008) that animation did not help compared to similar static visuals. However, this study is consistent with their suggestion that carefully designed animations toward the fostering of conceptual understanding are possible to have better performance than still visuals (Clark & Mayer, 2008).

## Conclusions

Although our self-directed learning experiment lasted for only seven class periods (30 minutes for each period) for three different groups of classes from only one school in Taiwan, the improvement made by the fifth graders in the computer-animation group is apparent and worthy of attention. Although generalizations of the present study may be limited, the results do provide some important contributions. Thus, further research studies should focus on long-term treatment and include a broader range of students in the studies.

First, the major contribution of the current study is that students' number sense can be promoted effectively by students themselves, if appropriate technology-enhanced learning activities are provided. The present study is different from an earlier study by Yang and Tsai (2010) that student's number sense could be promoted via teacher's instruction and the use of technology. These animated and number-sense-based activities emphasized understanding and potential applications of number sense and they allow students to self-study at any time through web-based or CD-based self-directed learning in order to help children develop number sense in an interactive and effective manner. Moreover, these animation number sense activities can be learned by students without the aid of teachers. It is also different from earlier non-technology-based studies (Griffin, 2004; Markovits & Sowder, 1994; Verschaffel et al., 2007; Yang & Wu, 2010) that children's number sense can be improved by appropriate "teacher-directed learning" in a traditional school mathematics classroom. Therefore, this study demonstrates three potentially helpful characteristics of the learning environment: number-sense-based, self-directed learning and animation-based.

Second, even though many studies and reports (Dick, 2007; Inamdar & Kulkarni, 2007; Lin, 2008; NCTM, 2000; Ruthven, 2007; Vulis & Small, 2007; Zbiek et al., 2007) have suggested that technology integrated into mathematics teaching and learning has a positive effect on children's mathematics performance, no study has focused on the effect of technology on number sense performance of elementary school children when self-directed learning via computer-based activities. This study initially shows that the appropriate use of technology can be a helpful learning tool for developing children's number sense.

Third, without teacher's teaching in class, children can develop number sense through self-directed learning. In cases in which children need more opportunities to learn and school cannot provide additional teaching time, these computer-animated number-sense activities could be a good substitute for developing their number sense. As earlier studies (Yang & Li, 2008; Yang & Wu, 2010) have suggested, an opportunity to learn plays a role in helping children develop number sense; providing children with ubiquitous access to these self-directed learning animated activities via the world-wide web would allow students to learn more about number sense whenever time is available to them. Consequently, these self-directed learning activities can serve as supplementary learning materials for helping children develop number sense at school or at home.

Finally, several factors caused the students learned more in the animated learning mode. First, animation activities through the use of computer can reinforce students' learning motivation. This supports the finding of previous studies that the integration of technology into mathematics classes has a positive effect on learning and motivation (Isiksal & Askar, 2005; Lin, 2008; Olkun, Altun, & Smith, 2005; Yang & Tsai, 2010). Second, the audio, words and visual display can simultaneously enhance the learning of number sense. This supports the finding of earlier studies that students may learn more from multimedia-based learning environment due to multimedia messages including words and visuals than what traditional teaching modes can do (Gegner et al., 2009; Mayer, 2003). Third, learning by doing is a good approach to promote students' learning. During manipulating the computerized activities, students will focus their attentions on the animation. Thus, it facilitates their number sense development.

In summary, the present study demonstrates that the integration of technology into number sense learning by appropriately utilizing multimedia allows learners to study visually, interactively, and effectively. However, the development of computer-animated number-sense activities is a complicated and time-consuming process. It requires the collaboration of mathematics educators who are specialists in number sense, computer programmers who are well experienced in writing such programs, and computer-animation designers. Therefore, to further the development of these animated activities, it may be necessary to have an educational center that can organize and design these computer- animated activities for self-directed learning.

### **Implications and practical applications**

The use of computer-animated, self-directed learning number-sense activities was successful in helping children enhance their number-sense performance and increase their uses of a higher order learning skill: number sense. The animated self-directed learning units for number sense can be made easily accessible at school or at home to any interested students through a web site on the world-wide web. Students can self-learn online without additional teachers' preparation or instruction time. These technology-based learning activities can also serve as supplemental learning materials in order to offer students extra opportunities to explore and develop number sense.

### **Limitations**

There are two main limitations of the current study. The first is that due to limited time and resources available, merely two intact classes were involved in the study. Therefore, generality of its results may need additional empirical evidences in other school settings. The second limitation relates to the extent to which the main aspects of number sense were investigated in the study. Only five aspects of number sense were covered, generality of the result to uncovered aspects of number sense may not be valid.

### **Acknowledgements**

This paper is a part of a research project supported by the National Science Council, Taiwan with grant no. NSC-100-2511-S-415-008 –MY3. Any opinions expressed here are those of the authors and do not necessarily reflect the views of the National Science Council, Taiwan. The authors also express their gratitude for the comments given by the Editor(s) and the two anonymous reviewers in assistance with the development of this article.

### **References**

- Bennison, A., & Goos, M. (2010). Learning to teach mathematics with technology: A survey of professional development needs, experiences and impacts. *Mathematics Education Research Journal*, 22(1), 31-56.
- Bobis, J. (2004). Number sense and the professional development of teachers. In A. McIntosh & L. Sparrow (Eds.), *Beyond written computation* (pp. 160–170). Perth, Western Australia: Mathematics, Science & Technology Education Centre, Edith Cowan University.
- Chan, H., Tsai, P., & Huang, T. Y. (2006). Web-based learning in a geometry course. *Educational Technology and Society*, 9(2), 133–140.
- Clark, R. C., & Mayer, R. E. (2008). Learning by viewing versus learning by doing: Evidence-based guidelines for principled learning environments. *Performance Improvement*, 47(9), 5-13.



- Dick, T. (2007). Keeping the faith: Fidelity in technological tools for mathematics education. In G. W. Blume & M. K. Heid (Eds.), *Research on technology and the teaching and learning of mathematics: Syntheses, cases, and perspectives. Vol. 2: Cases and perspectives* (pp. 333–339). Greenwich, CT: Information Age.
- Dunphy, E. (2007). The primary mathematics curriculum: Enhancing its potential for developing young children's number sense in the early years at school. *Irish Educational Studies* 26(1), 5–25.
- Forster, P. (2006). Assessing technology-based approaches for teaching and learning mathematics, *International Journal of Mathematical Education in Science and Technology*, 37(2), 145-164.
- Gegner, J. A., Mackay, D. H. J., & Mayer, R. E. (2009). Computer-supported aids to making sense of scientific articles: cognitive, motivational, and attitudinal effects. *Education Technology Research Development*, 57(1), 79-97.
- Goldin, G. A. (2002). Representation in mathematical learning and problem solving. In L. D. English (Ed.), *Handbook of international research in mathematics education* (pp.197–218). Mahwah, NJ: Erlbaum.
- Griffin, S. (2004). Building number sense with number words: A mathematics program for young children. *Early Childhood Research Quarterly*, 19(1), 173–180.
- Hudson, S., Kadan, S., Lavin, K., & Vasquez, T. (2010). *Improving basic math skills using technology*. (Unpublished master's these). Saint Xavier University, Chicago, Illinois.
- Inamdar, P., & Kulkarni, A. (2007). "Hole-in-the-wall" computer kiosks foster mathematics achievement—A comparative study. *Educational Technology & Society*, 10(2), 170–179.
- Isiksal, M., & Askar, P. (2005). The effect of spreadsheet and dynamic geometry software on the achievement and self-efficacy of 7th-grade students. *Educational Research*, 47(3), 333-350.
- Jordan, N. C., Glutting, J., & Ramineni, C. (2010). The importance of number sense to mathematics achievement in first and third grades. *Learning and Individual Differences*, 20(2), 82-88.
- Lin, C. Y. (2008). Beliefs about using technology in the mathematics classroom: Interviews with pre-service elementary teachers. *Eurasia Journal of Mathematics, Science & Technology Education*, 4(2), 135–142.
- Markovits, Z., & Sowder, J. T. (1994). Developing number sense: An intervention study in grade 7. *Journal for Research in Mathematics Education*, 25(1), 4–29.
- Mayer, R. (2003). The promise of multimedia learning: using the same instructional design methods across different media. *Learning and Instruction*, 13(2), 125-139.
- Mayer, R. E., Hegarty, M., Mayer, S., & Campbell, J. (2005). When static medial promote active learning: Annotated illustrations versus narrated animation in multimedia instruction. *Journal of Experimental Psychology: Applied*, 11(4), 256-265.
- McIntosh, A., Reys, B. J., Reys, R. E., Bana, J., & Farrel, B. (1997). *Number sense in school mathematics: Student performance in four countries*. Perth, Australia: Edith Cowan University.
- Menon, R. (2004a). Elementary school children's number sense [www document]. *International Journal for Mathematics Teaching and Learning*. Retrieved from <http://www.cimt.plymouth.ac.uk/journal/default.htm>
- Menon, R. (2004b). Pre-service teachers' number sense. *Focus on Learning Problems in Mathematics*, 26(2), 49–61.
- Ministry of Education in Singapore. (2009). *Secondary mathematics syllabus*. Singapore: Ministry of Education. Retrieved from: <http://www.moe.edu.sg/education/syllabuses/sciences/files/maths-secondary.pdf>
- Moyer, P. S., Niezgod, D., & Stanley, J. (2005). Young children's use of virtual manipulatives and other forms of mathematical representations. In W. J. Masalski & P. C. Elliot (Eds.), *Technology-supported mathematics learning environments. Sixty-seventh yearbook of the National Council of Teachers of Mathematics*. Reston, VA: NCTM.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.
- Olive, J., & Lobato, J. (2008). The learning of rational number concept of using technology. In M. K. Heid & G. W. Blume (Eds.), *Research on technology and the teaching and learning of mathematics: Research syntheses* (Vol. 1, pp. 1-53). Charlotte, NC: Information Age.
- Olkun, S., Altun, A., & Smith, G. (2005). Computers and 2D geometric learning of Turkish fourth and fifth graders. *British Journal of Educational Technology*, 36(2), 317-326.
- Philpot, T., Hall, R., Hubing, N., Flori, R., Oglesby, D., & Vikas, Y. (2003). Animates instructional media for stress transformations in a mechanics of materials course. *Computer Applications in Engineering*, 11(1), 40-52.
- Reimer, K., & Moyer, P. S. (2005). Third-graders learn about fractions using virtual manipulatives: A classroom study. *Journal of Computers in Mathematics and Science Teaching*, 24(1), 5–25.

- Reys, R. E., & Yang, D. C. (1998). Relationship between computational performance and number sense among sixth- and eighth-grade students in Taiwan. *Journal for Research in Mathematics Education*, 29(2), 225–237.
- Ruthven, K. (2007). Embedding new technologies in complex ongoing practices of school mathematics education. *International Journal for Technology in Mathematics Education* 13(4), 161–167.
- Stoffa, V. (2004). Modeling and simulation as a recognizing method in education. *Educational Media International*, 41(1), 51–58.
- Su, H. F., Marinas, C., & Furner, J. M. (2010). Connecting the numbers in the primary grades using an interactive tool. *Australian Primary mathematics Classroom*, 15(1), 25–28.
- Suh, J. M., & Moyer-Packenham, P. S. (2007). Developing students' representational fluency using virtual and physical algebra balances. *Journal of Computers in Mathematics and Science Teaching*, 26(2), 155–173.
- Taylor, M., Pountney, D., & Malabar, I. (2007). Animation as an aid for the teaching of mathematical concepts. *Journal of Further and Higher Education*, 31(3), 249–261.
- Tripathi, P. N. (2008). Developing mathematical understanding through multiple representations. *Mathematics Teaching in the Middle School*, 13(8), 438–445.
- Verschaffel, L., Greer, B., & De Corte, E. (2007). Whole number concepts and operations. In F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 557–628). Charlotte, NC: Information Age.
- Vulis, M., & Small, M. (2007). Why teaching business mathematics with technology might be very important in today's mathematics education. *International Journal for Technology in Mathematics Education*, 13, 212–213.
- Yang, D. C. (2010, January). *The design and application of web-based two-tier test for number sense*. Paper presented at the 8th Annual Hawaii International Conference on Education, Hawaii, U.S.A.
- Yang, D. C., & Li, M. N. (2008). An investigation of 3<sup>rd</sup> grade Taiwanese students' performance in number sense. *Educational Studies*, 34(5), 443–455.
- Yang, D. C., Reys, R. E., & Reys, B. J. (2009). Number sense strategies used by pre-service teachers in Taiwan. *International Journal of Science and Mathematics Education*, 7(2), 383–403.
- Yang, D. C., & Tsai, Y. F. (2010). Promoting sixth graders' number sense and learning attitudes via technology-based environment. *Educational Technology and Society*, 13(4), 112–125.
- Yang, D. C., & Wu, W. R. (2010). The study of number sense realistic activities integrated into third-grade math classes in Taiwan. *The Journal of Educational Research*, 103(6), 379–392.
- Zbiek, R. M., Heid, M. K., Blume, G. W., & Dick, T. M. (2007). Research on technology in mathematics education: A perspective of constructs. In F. K. Lester Jr., (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 1169–1207). Charlotte, NC: Information Age.

## Appendix 1

The self-learning process for two groups

Schedule for sequences	Class period	CAEG	PAG
L1: Decimal problems	1	Animated A1	PV A1
L2: Calculator & Estimation	1	Animated A2	PV A2
L3: Comparing fractions	1	Animated A3	PV A3
L4: $\frac{1}{2}$ as a benchmark	1	Animated A4	PV A4
L5: Estimation & Measurement	1	Animated A5	PV A5
L6: Fraction problems	1	Animated A6	PV A6
L7: Fraction & Decimal	1	Animated A7	PV A7

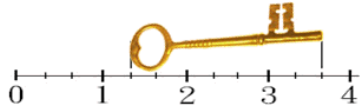
Note. Animated A1: Animated self-learning Activity1

PV A1: Paper version self-learning Activity1

## Appendix 2

An exemplified item

**Step1: Students are required to choose an answer.**

Question 16 / total questions of the test is 20	
Question	Which answer can best represent the length of the key on the number line below? 
Answer:	<input type="radio"/> $1\frac{1}{3} + 3\frac{2}{3}$
	<input type="radio"/> $1.1 + 3.2$
	<input type="radio"/> $3\frac{2}{3} - 1\frac{1}{3}$
	<input type="radio"/> $3.2 - 1.1$
<input type="button" value="Submit"/>	

**Step2: According to the answer, students are required to choose a reason for the answer!**

My reason is	
<input type="radio"/>	The key is sitting from $1\frac{1}{3}$ to $3\frac{2}{3}$ , so the length of the key is $1\frac{1}{3} + 3\frac{2}{3}$ cm.
<input type="radio"/>	You count to the right, which means the numbers are added together.
<input type="radio"/>	I'm guessing.
<input type="button" value="Submit"/>	

My reason is	
<input type="radio"/>	The key is sitting from 1.1 to 3.2, so the length of the key is $1.1 + 3.2$ cm.
<input type="radio"/>	The length of the key is close to 4, so it's about $1.1 + 3.2$ cm.
<input type="radio"/>	I'm guessing.
<input type="button" value="Submit"/>	

My reason is	
<input type="radio"/>	The key is sitting from $1\frac{1}{3}$ to $3\frac{2}{3}$ , so the length of the key is $3\frac{2}{3} - 1\frac{1}{3}$ cm.
<input type="radio"/>	Subtraction is needed to evaluate the length, so it's $3\frac{2}{3} - 1\frac{1}{3}$ cm.
<input type="radio"/>	I'm guessing.
<input type="button" value="Submit"/>	

My reason is	
<input type="radio"/>	The key is sitting from 1.1 to 3.2, so the length of the key is $3.2 - 1.1$ cm.
<input type="radio"/>	Subtraction is needed to evaluate the length, so it's $3.2 - 1.1$ cm.
<input type="radio"/>	I'm guessing.
<input type="button" value="Submit"/>	

## Goals, Motivation for, and Outcomes of Personal Learning through Networks: Results of a Tweetstorm

**Rory L. L. Sie<sup>1\*</sup>, Nino Patariaia<sup>2</sup>, Eleni Boursinou<sup>2</sup>, Kamakshi Rajagopal<sup>3</sup>, Anoush Margaryan<sup>2</sup>, Isobel Falconer<sup>2</sup>, Marlies Bitter-Rijkema<sup>3</sup>, Allison Littlejohn<sup>2</sup> and Peter B. Sloep<sup>3</sup>**

<sup>1</sup>LOOK Scientific Centre for Teacher Research, Open University of the Netherlands, Valkenburgerweg 177, 6419 AT Heerlen, The Netherlands // <sup>2</sup>The Caledonian Academy, Glasgow Caledonian University, 58 Port Dundas Road, Glasgow, G4 0HG, Scotland // <sup>3</sup>Centre for Learning Sciences and Technologies, Open University of the Netherlands, Valkenburgerweg 177, 6419 AT Heerlen, The Netherlands // rory.sie@ou.nl // nino.patariaia@gcu.ac.uk // eleni.boursinou@gcu.ac.uk // kamakshi.rajagopal@ou.nl // anoush.margaryan@gcu.ac.uk // isobel.falconer@gcu.ac.uk // marlies.bitter@ou.nl // allison.littlejohn@gcu.ac.uk // peter.sloep@ou.nl

\*Corresponding author

(Submitted April 17, 2012; Revised November 26, 2012; Accepted February 12, 2013)

### ABSTRACT

Recent developments in the use of social media for learning have posed serious challenges for learners. The information overload that these online social tools create has changed the way learners learn and from whom they learn. An investigation of learners' goals, motivations and expected outcomes when using a personal learning network is essential since there have been few empirical studies in the domain. Previous research focused on the factors that influence learning in virtual environments, but these studies were mainly conducted in an era in which online social media were not yet used for personal learning networks. The current paper reports findings of a study that examined factors impacting professional learning through networks. A personal learning network identification session and a brainstorm via Twitter (*Tweetstorm*) regarding goals, motivational factors and outcomes of learning through networks were conducted. The article concludes that seven factors play a pivotal role in professional learning through networks: sharing, motivation, perceived value of the network, feedback, personal learning, trust and support, peer characteristics and peer value. Also, different perspectives, motivation, social media and collaboration, reciprocity, intrinsic motivation, innovation, status and reputation and networking strategies play an important role in motivation. Future work should focus on investigating the interplay between factors that influence networked learning identified in this article.

### Keywords

Personal learning network, Social networks, Networking tools, Twitter, Networked learning

### Introduction

Nowadays, it has become a norm for professionals to develop themselves as part of their job. They attend courses and seminars, and sometimes their reward or even continuation of their professional activities depends on such *formal learning*. However, Boud and Hager (2012) note "... a move from ... development - to the input - the activity", whereas professional development chiefly takes place by "participation in the practice" (Boud & Hager, 2012), *informal learning* occurs during daily practice. Johnsson, Boud and Solomon (2012) report about cases in which interaction with others drives informal learning, by offering new perspectives from other working contexts that trigger new thinking in the professional learner's personal context, akin to the creative power of 'bridges' Burt (2004) reports. Nevertheless, unlike formal learning, informal learning is not rewarded nor recognised, mainly due to lacking information about how individuals learn through their network (networked learning) (Haythornthwaite & De Laat, 2010).

Monitoring social interactions (social media may be especially suited for this) can help identifying informal learning (e.g., De Laat, Lally, Lipponen, & Simons, 2007), but learners need to be motivated to exhibit the appropriate informal learning behaviour. To do so, one needs to investigate professional learners' strategy, or 'networking attitude' (Rajagopal, Joosten-ten Brinke, Van Bruggen, & Sloep, 2012). More explicitly, what motivates professional learners to engage in learning through their network? What do they learn (learning outcomes)? And why do they feel they learn (perception)?

We present findings from a new type of knowledge elicitation, the *Tweetstorm*. The Tweetstorm is an online, open brainstorm-like session via Twitter, a microblogging platform. The current article starts off with some background literature necessary for understanding the remaining part of the article. Then, two pivotal terms in this article are explained: Personal Learning Networks (PLNs) and the Tweetstorm. Next, we provide the methods that we employed to conduct the Tweetstorm. Subsequently, we present the results of the Tweetstorm, including the results of one of the phases of the Tweetstorm: the PLN identification session. Finally, we discuss the results and draw together our conclusions.

## Literature review

The question *whom we learn from* has a long history in educational research and several learning theories aim to capture the social process of learning. Bandura (1977) defines social learning as learning from others; modelling and imitating others' behaviour. Vygotsky (1978) underlines that learning, internalising behaviour, occurs by imitation; we learn from others by example. Wenger (1998) contends that learning is practice-driven; people share a common interest or practice. Learners influence and learn from one another as they engage in their "community of practice". Connectivism (Siemens, 2005), a theory that explicitly refers to learning with technology, claims that "learning is a process of connecting to specialized nodes or information resources". This includes learning from resources, or organizations that possess knowledge.

Dillenbourg (1999, p. 2) defines that we learn collaboratively by having "a situation in which two or more people learn or attempt to learn something together". Social media can assist in social, collaborative learning, but these should be tailored to learning practice (Väljataga & Fiedler, 2009). Four main types of activities are distinguished to describe *how* we learn at the workplace (Eraut, 2004): (1) participation in group activities, (2) working alongside others, (3) tackling challenging tasks, and (4) working with clients. The first, second and fourth point towards social, collaborative actions, which may be important for our understanding of personal, professional learning networks.

*What* we learn in the workplace ranges from task performance, awareness and understanding, personal development, teamwork, role performance, academic knowledge and skills, decision making and problem solving, to judgement (Eraut, 2004). Schank (1995) states that we internalise so-called *scripts* of consecutive actions when we learn by doing. This is similar to the social learning view of Bandura (1977), who claims that we learn from others by constructing a model of what others do and try to imitate this.

The reason *why* learners engage in learning through networks may be that they share a common interest or practice (Lave, 1991), are keen to exchange of ideas (Pirulli, 2009) and want to receive and provide support (Berlanga, Sloep, Kester, Brouns, Van Rosmalen, & Koper, 2008; Fetter, Berlanga, & Sloep, 2010; Van Rosmalen et al., 2007). They also call on each other when they have a problem to solve or knowledgeability to offer (Dekker & Kingma, 1999). Social support theories posit that network relationships offer both instrumental and emotional support to network members (Gerstick, Bartunek & Dutton, 2000). Instrumental relationships encompass resources such as professional advice, information, and expertise, whereas emotional relationships provide encouragement, friendship, support and ways of communicating information (Ibarra, 1993). Access to knowledge resources may guide learner engagement in learning networks (Hollingshead, Fulk, & Monge, 2002). Also, learner engagement is subject to the learner's interest (Billett, 2004).

The above literature review very much describes the 'how and why' of social learning, but to make a link between theory and praxis, we need to understand what drives one to learn via her network. We need to understand why one engages in learning via her personal learning network, in order for managers to attach proper rewards to informal learning actions. In other words, we need to bridge the gap between networked learning and the recognition, valuation and rewarding of it by managers of learners, for instance teachers (in the case of students) or line managers (in the case of organizational learning). Therefore, the current study attempts to explore how professionals utilise their networks and what motivates them to use their network for learning. Hence, this article investigates the following main research question:

*What makes professional learners learn through their network?*

To answer this question, we first need to lay out the definition of a personal learning network. In addition, we present the workflow and pros and cons of our elicitation method: the *Tweetstorm*.

## Personal learning networks and the Tweetstorm

There are mainly two approaches to personal learning networks: a top-down, and a bottom-up approach. From a top-down perspective, a learning network can be part of a collaborative learning *solution* that consists of introducing a networking environment for learners to become more motivated, or less isolated, by recommending them knowledgeable peers (Fetter, Berlanga, & Sloep, 2010; Sie et al., 2012). From a bottom-up perspective, connectivists such as Siemens (2005) contend that learning is a social *phenomenon*: learner interactions constitute a learning network. We refer to this act as *networked learning*. Networked learning can, for instance, be extracted from a personal learning environment (PLE) by monitoring who reacts to whom in a PLE's forum to guide interventions in the learning process (Corallo, De Maggio, Grippa, & Passiante, 2008). When a learning network is tailored to the individual, for instance to give learner-centric feedback or advice, it is called a personal learning network (PLN).

The PLN identification session we employ here as part of the Tweetstorm introduction is an instance of egocentric network elicitation. That is, we explore the relationships that are formed during networked learning. In detail, *egocentric networks* are networks from the perspective of an individual, a well-known data type in Social Network Analysis. The data collection method is simple: the researcher asks a participant for his or her contacts, and then draws a social network consisting of nodes (contacts) and edges (relationships). The 'networked learning'-relationships that we investigate are two-fold: we investigate from whom the participants learn, and which tools they use to do so.

The Tweetstorm is an online knowledge elicitation method. It uses the open, online character of Twitter combined with the strength of freely uttering statements from brainstorming. Also, it draws on expert knowledge to arrive at core statements, akin to concept mapping (Stoyanov & Kirschner, 2004) or the Delphi method. It consists of six phases: First, a *context* should be provided. This can be a short introduction of the subject of the Tweetstorm, or in this case, we additionally held a personal learning network identification session with the participants present at a workshop. Second, *questions are presented* through Twitter using a *hashtag*, a short, unique, textual code to denote the subject of a *tweet*, a Twitter message. Third, the participants *answer the questions* using their personal Twitter account and the designated hashtag. Fourth, the *tweets are aggregated* using the hashtag and an aggregation tool such as YourTwapperkeeper (<https://github.com/jobrieniii/yourTwapperKeeper>). Fifth, the *tweets are categorised* by a team of experts using a card sorting tool such as Websort ([www.websort.net](http://www.websort.net)). The experts put the tweets in categories that they can name themselves. Finally, the *categorisation is analysed*. The agreement between the experts' sorting is computed, without the need for experts to agree upon a category name. Based on this agreement, hierarchical cluster analysis software computes core clusters of tweets. Figure 1 shows an overview of the distinct phases.

The advantages of the Tweetstorm are numerous. First, it allows for easy display: search for a hashtag and all twitter statements, so-called *tweets*, containing the hashtag are returned. These can be easily displayed using a so-called *Twitterwall*, an aggregator specialised in the display of tweets of a particular hashtag. Second, and following the previous advantage, it allows for easy aggregation as participants' statements are already in digital format, and can be traced back to individual participants. Third, Twitter is a well-known medium, which allows for smooth setup of a Tweetstorm session. Fourth, the categorisation of tweets is performed online, which allows for experts to take part from any place, any time. Fifth, experts can take part anonymously, which prevents them from experiencing production blocking by other experts. That is, they can utter their statements without having to wait for their turn. Sixth, experts can work independently: experts can put tweets in a category they define themselves. In contrast to methods such as the Delphi (Hasson, Keeney, & McKenna, 2000), there is no need to reach consensus with other experts, which saves a lot of time. Finally, in contrast to many other idea generation or knowledge elicitation methods such as brainstorming (Osborn, 1954), Think-Pair-Share or Progressive Inquiry (Hakkarainen, 2003), the Tweetstorm is not limited solely to a selected group of invitees; due to the nature of Twitter, it is open to anyone who has knowledge of the session's hashtag.

The main drawback of the Tweetstorm method is that it is not entirely suited for in-depth discussion among participants. Participants are able to react to one another, and they are able to concur with one another by retweeting. However, tweets are limited by 140 characters, making it difficult to sufficiently make an argument.

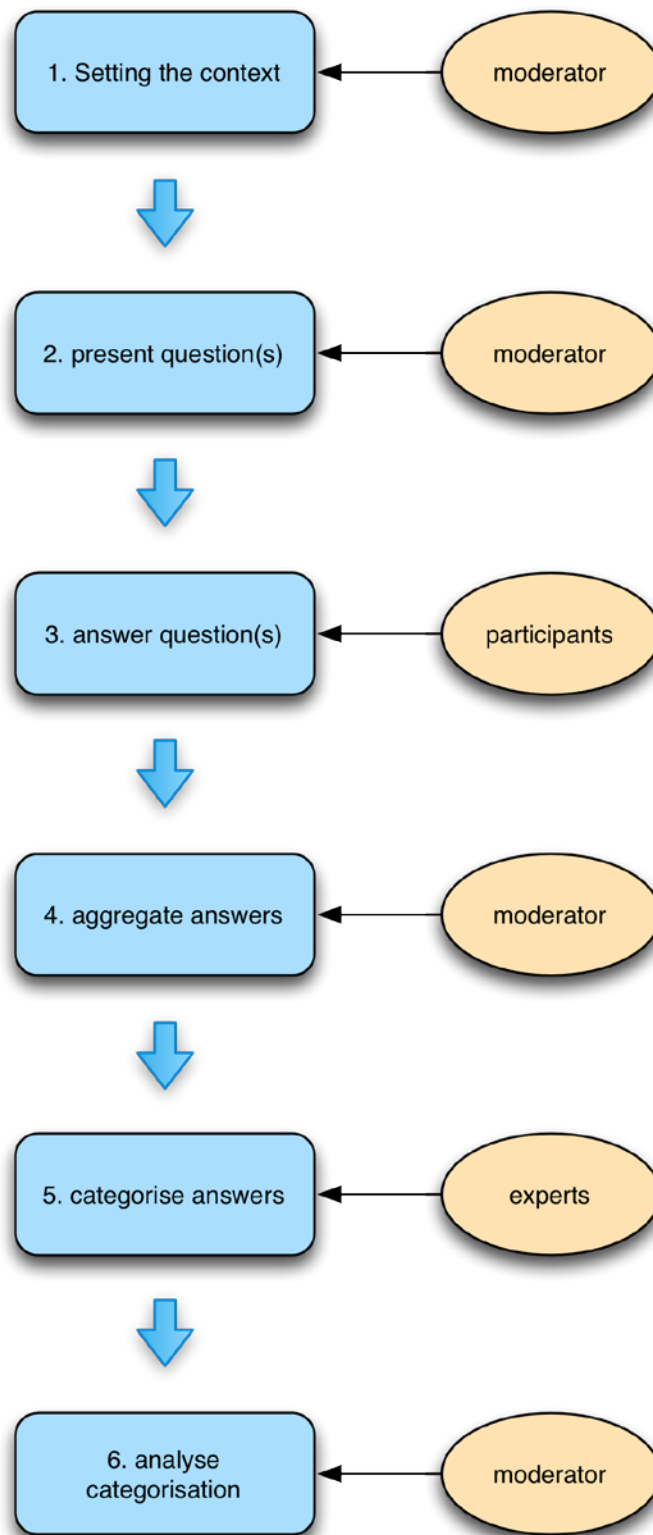


Figure 1. Overview of the Tweetstorm method



## Method

### Participants

#### *Setting the context: PLN identification session*

Participants were attendants of a workshop at the PLE conference 2011, which was announced before the start of the conference. Akin to Scardamalia and Bereiter's "quality circles" (1994), a conference allows scholars to learn from one another, to advance their knowledge more efficiently and effectively than going it alone. The six 'live' participants, attendants of the workshop that is to say, which excludes the online participants, were chiefly educational researchers with an interest in Personal Learning Environments. The participants' main characteristics are provided in Table 1. No inducement was offered for their participation.

*Table 1. Overview of the 'live' participants' main characteristics.*

ID	gender	age range	profession	discipline
1	m	35-44	PhD student	education
2	f	45-54	teacher	cultural and ethnic studies
3	m	35-44	professor	other
4	m	25-34	post-doc	education
5	m	25-34	PhD student	computer science
6	f	25-34	teacher	sociology

#### *Presenting and answering questions*

Due to the public nature of Twitter, the Tweetstorm was open to anyone who was interested and managed to spot it. The Tweetstorm was announced through the website of the PLE conference. A total of 31 participants actively engaged in it by *tweeting* (uttering 'tweets') or *retweeting* (forwarding tweets). This included the six participants that engaged in the antecedent PLN identification session. The use of Twitter meant that we could only identify participants by their Twitter username (quasi-anonymity). As indicated, passive, read-only participants ('lurkers') could also join the Tweetstorm. As Twitter does not allow for tracking of 'reads', lurkers could have (indirectly) influenced the Tweetstorm by discussing with active participants offline. No inducement was offered for participants' cooperation.


#### *Categorisation and analysis*

We invited a group of experts to participate in a sorting experiment to independently categorise the statements that were extracted from the tweets. Since the statements were about learning in networks, 34 experts from affiliated universities, researchers in the educational domain, were invited via email, of which nine responded positively. Their occupation varied from PhD student to associate professor. Again, no inducement was offered for their help.

### Materials

#### *PLN identification session*

A custom-built online environment (PLN identification tool) was used in which participants could register themselves and identify the contacts in their PLN (Figure 2). The environment was accessible through the Internet URL 145.20.132.20/rse/test/page/PLE. For ease of use, the URL given to the participants was shortened using an online service called Bit.ly. The environment was tested during a pilot session at Glasgow Caledonian University. Five participants, all educational researchers, tested the environment and were given the opportunity to 1) reflect on clarity and usefulness of the questions, and 2) to provide suggestions for improvement. As a result, the survey instruments and questions were refined prior to the actual session. Although some of the answer options that were added seem to overlap with the existing ones, the participants of the pilot felt these needed to be added. For instance, 'external colleague' and 'research collaborator' may have overlap in meaning.



## CoCooN: Coalitions for Cooperation Networks

[Logout](#)  
[PLN contacts form](#)

### In your daily professional life, who do you learn from?

Please enter your contact's details below.

First name:

Last name:

What is your relationship to the other person?

- ☐ internal colleague
- ☐ external colleague
- ☐ friend
- ☐ family
- ☐ friend of a friend
- ☐ project member
- ☐ random
- ☐ PhD student
- ☐ flatmate
- ☐ Supervisor
- ☐ Previous Lecturer
- ☐ Previous Supervisor
- ☐ Organisation
- ☐ research collaborator, co-author
- ☐ Line manager

Other: separate by commas

Is it a weak or a strong tie?

Why do you feel you learn from that person?

What tool/technology do you use to connect to that person?

- ☐ LinkedIn
- ☐ Facebook
- ☐ Twitter
- ☐ Email
- ☐ Face to face
- ☐ Phone/text
- ☐ Blog
- ☐ Skype
- ☐ Google docs
- ☐ forums
- ☐ Wiki
- ☐ Podcasts
- ☐ Delicious
- ☐ Mendeley
- ☐ Youtube

Other: separate by commas

Your current contacts:

Figure 2. Screenshot of the PLN identification tool.

### *Presenting and answering questions and aggregation*

The YourTwrapperkeeper aggregation tool, a so-called *twitterwall* (Figure 3) was shown at the workshop venue to present an overview of all tweets with the hashtag #plntweet. Participants could easily view what questions were asked, and what other participants answered. Also, the YourTwrapperkeeper tool allowed for easy aggregation of the tweets for analysis.



Figure 3. Part of the YourTwapperkeeper twitterwall used at the workshop venue.

### *Categorising and analysing categories*

The statements that resulted from the tweets were categorised by experts using Websort ([www.websort.net](http://www.websort.net)), which is designed to do card sorting experiments and corresponding data analysis. Having the statements in digital form allows for card sorting online. WebSort provides a number of data aggregation (e.g. items vs. items, items vs. categories) and visualisation methods (e.g. tree structure, tables). Participants were not able to see each other's categorisations. Also, the categorisation did not impose any time-constraints.

The multidendrograms software package (Fernández & Gómez, 2008) was used to perform agglomerative hierarchical cluster analysis (AHCA) with complete linkage (Defays, 1977) to find core clusters of statements. AHCA starts with  $n$  statements in  $n$  distinct clusters. In subsequent iterations, high similarity clusters are merged, until the appropriate number of clusters is reached. The main advantage of such an iterative, hierarchical method is that one can see how the assignment of statements to clusters occurs.

## **Procedure**

### *PLN identification session*

The PLN identification session lasted 45 minutes in total. At registration, participants of the PLN identification session described their profile in terms of their age range, gender, occupation, discipline and work experience. The main advantage of providing and keeping login credentials is that participants can be asked to identify contacts at a later point in time (repeated measure), to see how their network and perception of this network evolves. Afterwards, participants could add contacts that they learn from through the PLN contacts form. For each contact, the participants had to answer the following questions:

- What is your relationship to the other person?

- Is it a weak or a strong tie?
- Why do you feel you learn from that person?
- What tool/technology do you use to connect to that person?

Participants could edit or delete the contacts that they entered (bottom of Figure 2; actual entries are left out for privacy reasons). Although participants were asked to identify their learning contacts, the relationships between contacts and contacts' characteristics were not analysed. Using SPSS statistical software version 18, we calculated averages per type of contact and tool that learners used to connect to their learning contacts.

### *Tweetstorm*

During the Tweetstorm, which lasted 45 minutes in total, participants were asked to tweet their personal opinion using the #plntweet hashtag. In that way, the moderators could aggregate all tweets after the Tweetstorm had ended. The moderators (three) tried to trigger participants by posting three main questions about PLNs to Twitter using the #plntweet hashtag:

- What motivates you to engage/learn through your network?
- Why do you feel you learn from your peers?
- What do you learn from your network?

### *Statement sorting*

The tweets were aggregated and split up into smaller pieces of information, as most of the tweets addressed multiple questions at once. One tweet could answer both the question on motivation (what motivates the learner) and on the content (what is learned) of learning. As the researchers posted (tweeted) the triggering questions separately, it was not expected that participants would answer multiple questions in a single tweet. Moreover, some of the answers contained distinct parts that could possibly be interpreted and categorised differently from each other. For example, one part of the answer could be about feedback, whereas another part could be about inspiration. After splitting up these tweets into separate statements, we uploaded these in the Websort.net environment. Following this, we asked the experts to categorise the statements. To prevent researcher bias, no pre-defined categories were provided. Experts could define and name categories themselves.

We used the Websort environment to export the sorting data to an *item-item* similarity matrix. This matrix is too large to be reported here in full, however it is available on [http://www.open.ou.nl/rse/Rory\\_Sie/Downloads.html](http://www.open.ou.nl/rse/Rory_Sie/Downloads.html). Finally, AHCA with complete linkage was performed to find core clusters of statements.

### *Limitations*

The results of the PLN identification session were difficult to analyse by character, as they consisted of some multiple response questions, which means that a contact could be a research collaborator and an external colleague at the same time. Also, the response rate was relatively low. Further investigation with a larger group of participants is needed to allow more robust PLN identification. A further study with a larger group of participants would also allow us to aggregate the egocentric networks and compare the participants' view of their network to existing learning networks of which they are a part.

A further limitation of this study was that participants were mostly researchers already with a shared interest as evidenced by their attendance at this particular conference. Thus, the answers are likely to be in line with this type of profession. Future research should try to focus on participant groups beyond academia, in order to arrive at more general findings. Finally, the Tweetstorm results may have been influenced by the fact that it was a brainstorm that took place via Twitter. The participants were inexperienced with such type of elicitation, which may have had its influence on the way participants expressed their statements.

## Results

### PLN identification session

#### *Whom do participants learn from?*

Fifteen types of connections and fifteen different tools for communication were identified in the answers by the participants of the introductory session (Figure 4). From the six participants, one participant had named only five contacts. The rest had identified more than ten contacts, ranging from ten to twenty-four. In total, 261 contacts were identified. The participants could be connected to the same peer by more than one type of connection or tool. For example, a research collaborator could also be the participant's friend and use face-to-face as well as email communication. The findings revealed that the most common type of relationship in a learning network was research collaborator, friend and external colleague. 40% of research collaborators were at the same time friends. Following in order of meaningful connections were internal colleagues and supervisors.

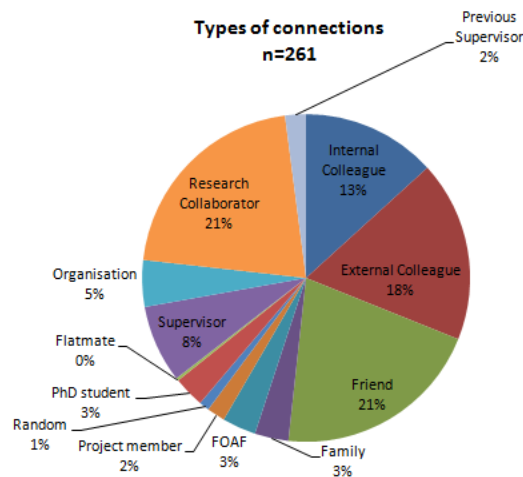


Figure 4. Whom do people learn from?

#### *What tools do they use?*

In total, thirteen out of fifteen distinct tools were selected by participants (Figure 5). The tools used most commonly were Twitter (18%, per participant:  $M=.68$ ,  $SD=.47$ ), email (19%, per participant:  $M=.65$ ,  $SD=.48$ ) and face-to-face communication (18%, per participant:  $M=.65$ ,  $SD=.48$ ). Although the social bookmarking tools Delicious and Wikis were an option, they were not mentioned.

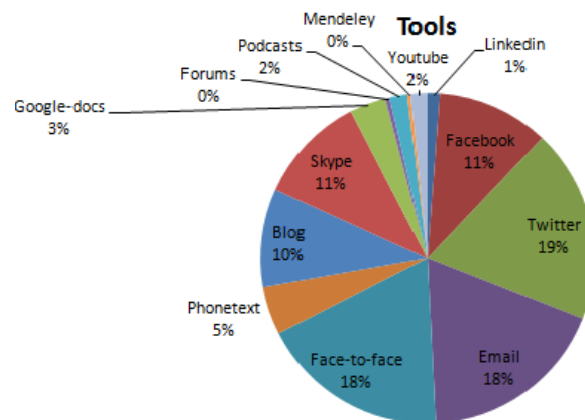


Figure 5. Tools used to learn from peers.

## Tweetstorm

Participants posted a total of 139 tweets ( $M = 4.48$ ;  $SD = 6.28$ ) (38 retweets) with the requested #plntweet hashtag. Sorting of the tweets entailed that we had to remove retweets, triggering questions, and split up tweets with multiple statements in them. A total of 83 statements were extracted from the Tweetstorm (see [http://www.open.ou.nl/rse/Rory\\_Sie/Downloads.html](http://www.open.ou.nl/rse/Rory_Sie/Downloads.html)).

## Statement sorting

There was no time-constraint set for the sorting exercise. Experts spent 51 minutes on average sorting the statements ( $SD = 35$ ). Nearly no overlap between experts' category names was found (Appendix A). The reason for this is clear and expected; the experts could define the category names themselves. Nevertheless, latent agreement between experts could be detected using agglomerative hierarchical analysis (Figure 6). On the lowest level, seven core clusters of statements can be found. When taking a closer look at the widest rectangle toward the bottom left, one remarkably notices that this cluster actually comprises eight subclusters. In an attempt to merge the single-statement cluster (a70) with, or assign to other clusters, the complete linkage clustering algorithm merged it with seven other clusters.

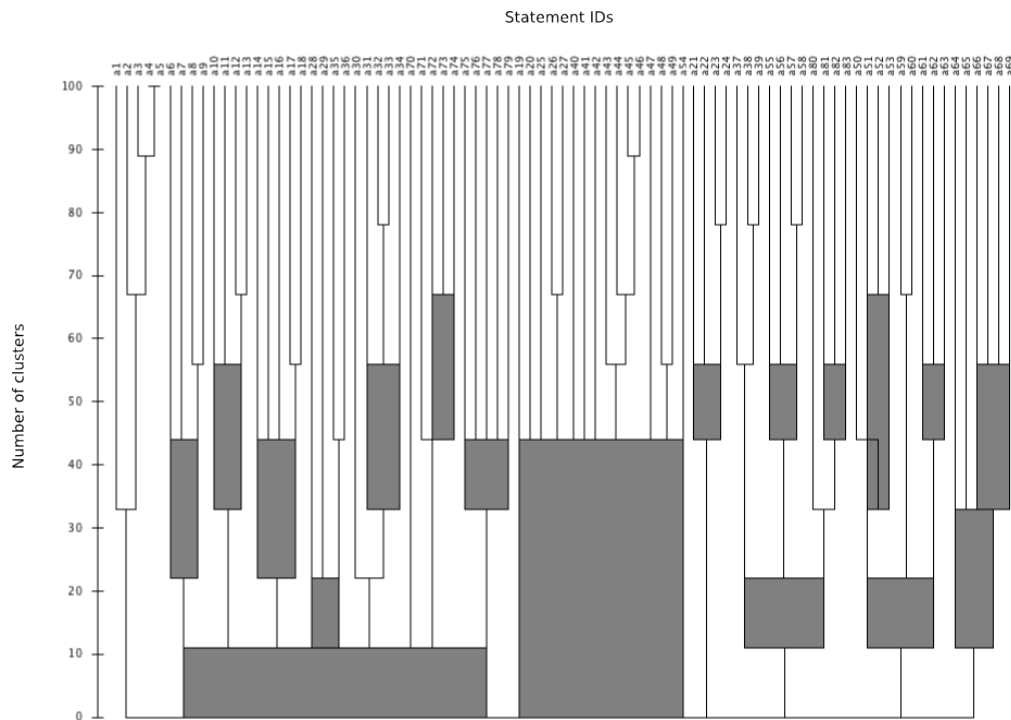


Figure 6. Results of hierarchical cluster analysis. Statements are coded, and can be found at [http://www.open.ou.nl/rse/Rory\\_Sie/Downloads.html](http://www.open.ou.nl/rse/Rory_Sie/Downloads.html).

Table 2 shows the number of statements per cluster, and typical statements for that cluster. Cluster 2 is the cluster corresponding to the abovementioned cluster that actually contains eight subclusters. For clarification, an overview of these clusters is given in Figure 7.

Table 2. Statements per cluster at the level of seven core clusters.

#	name	no. of statements	example
1	Sharing	5	“sharing is key”
2	Motivation	32	“Learning with others is more rewarding and rich than on your own”
2.1	Different	4	“Learn from your peers - Views I hadn't considered, opinions I disagree

	perspectives		with, ideas that inspire me"
2.2	Motivation	4	"For me, learning through my network is the most fun way of learning"
2.3	Social media and collaboration	5	"Twitter is a fine balance between the personal and the social. No-one learns in a vacuum, but we all learn uniquely"
2.4	reciprocity	4	"Conversation is 2-way. I can give to my network as well as take from it"
2.5	intrinsic motivation	5	"I use my PLN because of the autonomy it provides me"
2.6	Innovation	1	"By results collaboratively achieved - new methods under construction e.g. by MOOC ing. Old scales don't work."
2.7	status and reputation	4	"Not everyone has equal status in my PLN"
2.8	networking strategies	5	"My PLN allows me to connect to new people, communities and artefacts"
3	Perceived value of the network	16	"Finding out about latest research"
4	Feedback	4	"Feedback on thoughts and ideas"; "Instantaneous feedback, news, useful links, arguments and opinions"
5	Personal learning	11	"Using my network to find information and learn is the most effective and fast way to get the things I need"
6	Trust and support	9	"Ask for help and they will engage and help me"; "I can also discuss some of the concerns and insecurities I have within a peer group informally"
7	Peer characteristics and value	6	"Members of my PLN are very intelligent, inspirational, insightful and innovative"; "The people I learn from are passionate, critical and informed. They are my role models learners [sic] in this digital age"

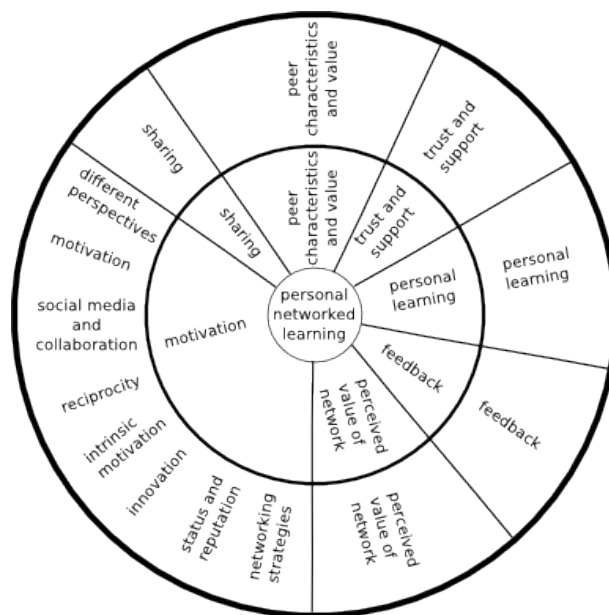


Figure 7. Seven core clusters and their fourteen subclusters.

## Discussion

The PLN identification session, which focused on identification of egocentric networks, revealed some interesting findings. The results show that participants learn mainly from research collaborators, friends and external colleagues. This is partly inline with research by Hoffman and Vance (2005), who found that individuals learn IT skills mainly by themselves, from family, or from friends. Though, Hoffman and Vance report that instant messaging was, at that

time, “a relatively new technology task”. Now that instant messaging and social media have integrated in daily lives of people, it may well be that a shift has taken place from learning mainly from offline contacts to learning from online contacts as well. Our findings of which medium was used to learn from peers support this: Learners mainly used face-to-face, email and Twitter as main modes of communication. Twitter did not yet exist back in 2005, which may explain the shift in whom people learn from.

The Tweetstorm and the corresponding agglomerative hierarchical cluster analysis resulted in a core set of seven clusters and a core set of fourteen clusters, of which eight subclusters were part of the larger cluster motivation. At the level of the seven clusters, the cluster ‘sharing’ is consistent with research by Olson, Grudin and Horvitz (2004, p. 1) who state “*Information sharing is of immense value in the workplace because it reduces duplication of effort, and sits at the foundations of collaboration*”. On the other hand, Fogel and Nehmad (2009) report that a majority of men and women included a picture of themselves in their profile, but did not share their phone number and home address. Thus, people only share personal information to a limited extent. Though, Swan (2002) stresses the importance of interaction for teaching and learning in a network.

The opposing views of Olson et al. and Fogel and Nehmad support that trust (cluster 6) is important in a personal learning network, but also call for a balance between information sharing and trust. Furthermore, the importance of trust and support for learning is partly supported by Lankau and Scandura (2002), who contend that there exists a positive relationship between vocational support (mentoring in the workplace) and personal learning. In that same study, it was found that roles are an important indicator for skill development, which supports our findings that ‘peer characteristics and value’ play a significant role in personal learning networks.

Ames and Archer (1988, p. 264) report that “*a mastery goal orientation may foster a way of thinking that is necessary to sustain student involvement in learning as well as increase the likelihood that students will pursue tasks that foster increments in learning*”. This is in line with our cluster motivation and its subclusters motivation and intrinsic motivation. Though, the concept of mastery or control itself was not mentioned in any of the statements. Networking strategies, a subcluster of motivation, is consistent with research by Zimmerman, Bandura, & Martinez-pons (1992), who conclude that learning strategies play an important role in academic self-motivation. More specifically, the statements in the cluster networking strategies point towards connecting to the right peers in the network. In research about creativity and innovation it is found that connecting to the right peers in a network leads to more creativity (Burt, 2004; Kratzer & Lettl, 2008). This ability to combine perspectives is part of the key competences necessary for lifelong learning (European Commission, 2010).

## Conclusions

This paper presents findings of an exploratory study using an innovative elicitation technique called the Tweetstorm; the study aims to discover how learners perceive their personal learning in a network. Especially now that learning is increasingly using online, social technologies, a new study was needed to investigate the question at hand. Findings suggest that these professional learners, scholars, learn mainly from research collaborators, friends and external colleagues. This is supported by the social tools that these learners use: mostly email and Twitter. However, a larger sample is needed to draw definite conclusions.

The Tweetstorm results are mainly inline with current literature, which indicates that the motivation for networked learning, as well as why and what learners learn, has not changed with the increase of social media use. For example, a learner may be motivated through reciprocity (Kogut, 1989; Song, 2009) in the network (Aviv & Ravid, 2005). Learners want to have a *quid pro quo*; something in return for what they share in the network. For instance, in exchange for their participation and knowledge sharing, networked learners expect to receive feedback from other participants in the network. Furthermore, a personal learning network should keep a balance between an appropriate amount of information sharing and interaction in the network and a trustworthy and supportive entourage (Rusman, Van Bruggen, Cörvers, Sloep, & Koper, 2009). Future work should therefore focus on the interplay between factors that influence the interaction between networked learners. This entails further investigation about the importance for each factor for networked learning.



## Acknowledgments

We thank the respondents for their participation in the study. Furthermore, we are indebted to the experts Francis Brouns, Pia Fontana, Henry Hermans, Jo Boon, Teresa Guasch, Arnoud Evers, Bieke Schreurs, Ellen Rusman, Colin Milligan and Andrea Klaijsen for categorizing the tweets. We thank the reviewers for their insightful comments. Finally, we thank the organisers of the PLE conference 2011 for accepting our proposal for the workshop through which we collected the data for this study. We are indebted to Elisabeth Uijtenbroek for proofreading and helping improve this article.

## References

- Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology*, 80(3), 260–267. doi:10.1037//0022-0663.80.3.260
- Aviv, R., & Ravid, G. (2005). Reciprocity analysis of online learning networks. *Journal of Asynchronous Learning Networks*, 9(4), 1-10.
- Bandura, A. (1977). *Social learning theory*. New York, NY, USA: General Learning Press.
- Berlanga, A. J., Sloep, P. B., Kester, L., Brouns, F., Van Rosmalen, P., & Koper, R. (2008). Ad hoc transient communities: towards fostering knowledge sharing in learning networks. *International Journal of Learning Technology*, 3(4), 443-458. doi:10.1504/IJLT.2008.019378
- Billett, S. (2004). Workplace participatory practices: Conceptualising workplaces as learning environments. *Journal of Workplace Learning*, 16(6), 312–324. doi:10.1108/13665620410550295
- Boud, D., & Hager, P. (2012). Re-thinking continuing professional development through changing metaphors and location in professional practices. *Studies in Continuing Education*, 34(1), 17–30. doi: 10.1080/0158037X.2011.608656
- Burt, R. S. (2004). Structural holes and good ideas. *American Journal of Sociology*, 110(2), 349-399.
- Corallo, A., De Maggio, M., Grippa, F., & Passiante, G. (2008). Evolving mechanisms of virtual learning communities : Lessons learned from a case in higher education. *Emerging Technologies and Information Systems for the Knowledge Society*, 5288, 304–313.
- Defays, D. (1977). An efficient algorithm for a complete link method. *The Computer Journal*, 20(4), 364-366.
- Dekker, J. M., & Kingma, J. (1999). 'Managen' van leernetwerken. In J. M. Dekker (Ed.), *Netwerkend Leren (Opleiders in Organisaties: Capita Selecta vol 38)*. Deventer, The Netherlands: Kluwer.
- De Laat, M., Lally, V., Lipponen, L., & Simons, R.-J. (2007). Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for Social Network Analysis. *International Journal of Computer-Supported Collaborative Learning*, 2(1), 87–103. doi: 10.1007/s11412-007-9006-4
- Dillenbourg, P. (1999). What do you mean by “collaborative learning”? In P. Dillenbourg (Ed.), *Collaborative-learning: Cognitive and Computational Approaches* (Vol. 1, pp. 1-15). Oxford: Elsevier.
- Eraut, M. (2004). Informal learning in the workplace. *Studies in Continuing Education*, 26(2), 247-273. doi:10.1080/158037042000225245
- European Commission. (2010). *Education and training 2010 work programme*. Retrieved from [http://ec.europa.eu/education/lifelong-learning-policy/doc/clusters/peer07\\_en.pdf](http://ec.europa.eu/education/lifelong-learning-policy/doc/clusters/peer07_en.pdf)
- Fernández, A., & Gómez, S. (2008). Solving non-uniqueness in agglomerative hierarchical clustering using multidendrograms. *Journal of Classification*, 25(1), 43-65. doi: 10.1007/s00357-008-
- Fetter, S., Berlanga, A. J., & Sloep, P. (2010). Fostering social capital in a learning network: Laying the groundwork for a peer-support service. *International Journal of Learning Technology*, 5(4), 388-400.
- Fogel, J., & Nehmad, E. (2009). Internet social network communities: Risk taking, trust, and privacy concerns. *Computers in Human Behavior*, 25(1), 153–160. doi: 10.1016/j.chb.2008.08.006
- Gerstick, C. J. G., Bartunek, J. M., & Dutton, J. E. (2000). Learning from academia: The importance of relationships in professional life. *The Academy of Management Journal*, 43(6), 1026-1044. doi: 10.2307/1556333

- Hakkarainen, K. A. I. (2003). Emergence of progressive-inquiry culture in computer-supported collaborative learning. *Learning Environments Research*, 6(2), 199–220.
- Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of advanced nursing*, 32(4), 1008–15.
- Haythornthwaite, C., & De Laat, M. (2010). Social networks and learning networks: Using social network perspectives to understand social learning. In L. Dirckinck-Holmfeld, V. Hodgson, C. Jones, M. de Laat, D. McConnell, & T. Ryberg (Eds.), *Proceedings of the 7th International Conference on Networked Learning* (pp. 183–190). Aalborg, Denmark: Lancaster University.
- Hoffman, M. E., & Vance, D. R. (2005). Computer literacy: What students know and from whom they learned it. In W. Dann, T. Naps, P. Tymann, & D. Baldwin (Eds.), *Proceedings of the 36th SIGCSE Technical Symposium on Computer Science Education* (pp. 356–360). St. Louis, MO: ACM.
- Hollingshead, A. B., Fulk, J., & Monge, P. (2002). Fostering intranet knowledge sharing: An integration of transactive memory and public goods approaches. In S. Keisler & P. Hines (Eds.), *Distributed work: New Research on Working Across Distance Using Technology* (pp. 335–355). Cambridge, MA: MIT Press.
- Ibarra, H. (1993). Personal networks of women and minorities in management: A conceptual framework. *Academy of Management Review*, 18(1), 56–87.
- Johnsson, M. C., Boud, D., & Solomon, N. (2012). Learning in-between, across and beyond workplace boundaries. *International Journal of Human Resources Development and Management*, 12(1–2), 61–76.
- Kogut, B. (1989). The stability of joint ventures: Reciprocity and competitive rivalry. *Journal of Industrial Economics*, 38(2), 183–198.
- Kratzer, J., & Lettl, C. (2008). A social network perspective of lead users and creativity: An empirical study among children. *Creativity and Innovation Management*, 17(1), 26–36.
- Lankau, M. J., & Scandura, T. A. (2002). An investigation of personal learning in mentoring relationships: Content, antecedents, and consequences. *Academy of Management Journal*, 45(4), 779–790.
- Lave, J. (1991). Situating Learning in Communities of Practice. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 63–82). Washington, DC, USA: American Psychological Association. doi:10.1037/10096-003
- Olson, J. S., Grudin, J., & Horvitz, E. (2004). Toward understanding preferences for sharing and privacy. *Public Policy. Proceedings of the Extended Abstracts on Human Factors in Computing Systems* (pp. 1985–1988). Oregon, US: ACM.
- Osborn, A. F. (1954). *Applied imagination*. New York, NY, USA: Scribner.
- Pirolli, P. (2009). An elementary social information foraging model. *Proceedings of the 27th international conference on Human factors in computing systems (CHI'09)* (pp. 605–614). Boston, MA, USA.
- Rajagopal, K., Joosten-ten Brinke, D., Van Bruggen, J., & Sloep, P.B. (2012). Understanding personal learning networks: their structure, content and the networking skills needed. *First Monday*, 17(1), 1–29. doi:10.1145/1518701.1518795
- Rusman, E., Bruggen, J. V., Cörvers, R., Sloep, P., & Koper, R. (2009). From pattern to practice: evaluation of a design pattern fostering trust in virtual teams. *Computers in Human Behavior*, 25(5), 1010–1019.
- Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *Journal of the Learning Sciences*, 3(3), 265–283.
- Schank, R. C. (1995). *What we learn when we learn by doing* (Technical Report No. 60). Evanston, IL, USA: The Institute for the Learning Sciences, Northwestern University.
- Sie, R. L. L., Berlanga, A. J., Rajagopal, K., Pannekeet, K., Drachsler, H., Fazeli, S., & Sloep, P.B. (2012). Social tools for networked learning: Current and future research directions. In V. Hodgson et al. (Eds.), *Proceedings of the Eighth International Conference on Networked Learning 2012* (pp. 312–319). Maastricht, The Netherlands: Lancaster University.
- Siemens, G. (2005). Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning*, 2(1), 3–10.
- Song, F. (2009). Intergroup trust and reciprocity in strategic interactions: Effects of group decision-making mechanisms. *Organizational Behavior and Human Decision Processes*, 108(1), 164–173.
- Stoyanov, S., & Kirchner, P. (2004). Expert concept mapping method for defining the characteristics of adaptive e-learning: ALFANET project case. *Educational Technology Research and Development*, 52(2), 41–54.

- Swan, K. (2002). Building learning communities in online courses : The importance of interaction. *Education, Communication & Information*, 2(1), 23-49.
- Väljataga, T., & Fiedler, S. (2009). Supporting students to self-direct intentional learning projects with social media. *Educational Technology & Society*, 12(3), 58-69.
- Van Rosmalen, P., Sloep, P., Kester, L., Brouns, F., de Croock, M., Pannekeet, K., & Koper, R. (2007). A learner support model based on peer tutor selection. *Journal of Computer Assisted Learning*, 24(1), 74-86.
- Vygotsky, L. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
- Wenger, E. (1998). Communities of practice: Learning as a social system. *The Systems Thinker*, 9(5), 5.
- Zimmerman, B. J., Bandura, A., & Martinez-pons, M. (1992). Self-motivation for academic attainment : The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal*, 29(3), 663-676.

## Appendix A

### Categorisation by experts

Category	Experts	Total items	Unique items	Agreement
(Learning) benefits	1	24	24	1
Advantages	1	16	16	1
And take	1	9	9	1
Autonomy	1	1	1	1
Balance between give and take. Economic/rational approach	1	3	3	1
Based on a negative attitude	1	2	2	1
Characteristics of PLN	1	13	13	1
Characteristics/features of a network	1	12	12	1
Collaboration and community	1	5	5	1
Collaborative learning (with peers)	1	8	8	1
community identity, less relevant for me	1	4	4	1
Competences needed to be part of a Network	1	4	4	1
creation of a community of learners	1	13	13	1
definition of a network	1	16	16	1
Different conceptions of a PNL	1	15	15	1
Difficulties/problems	1	2	2	1
diversity	1	3	3	1
don't agree	1	5	5	1
effectiveness	1	3	3	1
efficiency	2	2	1	1
Expectatives	1	11	11	1
experiences	1	6	6	1
Feedback	1	4	4	1
Fun, happiness	1	3	3	1
fun, passion	1	13	13	1
General benefits of learning in a network: acquiring reputation/status based on quality of ideas	1	6	6	1
General benefits of learning in a network: efficiency/easiness/efficacy	1	2	2	1
General benefits of learning in a network: motivation/inspiration/passion	1	5	5	1
General benefits of learning in a network: quality/diversity/newness of ideas/perspectives	1	5	5	1
General benefits of learning in a network: role-modelling/examples/(common)reference framework	1	11	11	1
General benefits of learning in a network: supporting each other	1	6	6	1
General benefits of learning in a network: tailored to personal learning needs	1	4	4	1
Getting the world inside	1	12	12	1
Getting your world outside	1	5	5	1
Give	1	6	6	1
Goals	1	2	2	1
hmmm	1	1	1	1
hype	1	7	7	1
I don't understand :(	1	2	2	1
Ideas, information, inspiration and opinions	1	19	19	1
innovation	1	1	1	1
instruction	1	8	8	1
Interaction and support	1	6	6	1
interpretations	1	1	1	1
intrinsic motivation	1	8	8	1
intrinsic motivation from connecting to people	1	8	8	1
Knowledge, expertise	1	10	10	1
learning by interactions	1	23	23	1

learning goal	1	5	5	1
learning in networks	1	2	2	1
learning mainly as social learning=social exchange	1	13	13	1
learning to learn	1	9	9	1
learning=individual benefit receiving	1	39	39	1
limitations	1	1	1	1
maintain relations	1	3	3	1
make work interesting and inspirational	1	27	27	1
Misconceptions	1	5	5	1
models and expertise	1	6	6	1
Motivation	2	21	14	0.75
motivation: give and take	1	1	1	1
Motivations to be part of a Network	1	9	9	1
opinions	1	3	3	1
passion	1	2	2	1
pathetic statements	1	3	3	1
peers	1	3	3	1
People in My Network	1	13	13	1
perceived support by the network	1	12	12	1
Personal development	1	2	2	1
personal drive	1	7	7	1
personal gains by the network of learners	1	29	29	1
Personal learning due to participation in a network	1	12	12	1
platitudes	1	2	2	1
Problem solving and ask for help	1	6	6	1
Real-time interaction	1	3	3	1
Reasons for PLN	1	12	12	1
Reasons of learning (general)	1	3	3	1
Reflection and feedback often with peers	1	11	11	1
relying on others	1	14	14	1
reputation	2	6	5	0.6
resources	1	10	10	1
Roles	1	3	3	1
self-confidence	1	1	1	1
sharing	4	36	23	0.39
Social, informal interaction	1	5	5	1
Status	2	11	7	0.79
Stay in touch, connecting	1	5	5	1
Stay up-to-date	1	4	4	1
Support	1	3	3	1
trust, secure	1	3	3	1
Twitter	1	2	2	1
use network strategically	1	19	19	1
use of ICT	1	6	6	1

## A Comparative Evaluation of E-learning and Traditional Pedagogical Process Elements

**Damjan Vavpotič<sup>1\*</sup>, Boštjan Žvanut<sup>2</sup> and Irena Trobec<sup>2</sup>**

<sup>1</sup>University of Ljubljana, Faculty of Computer and Information Science, Tržaška 25, 1000 Ljubljana, Slovenia //

<sup>2</sup>University of Primorska, Faculty of Health Sciences, Polje 42, 6310 Izola, Slovenia // damjan.vavpotic@fri.uni-lj.si  
// bostjan.zvanut@fvz.upr.si // irena.trobec@fvz.upr.si

\*Corresponding author

(Submitted April 03, 2012; Revised November 08, 2012; Accepted December 19, 2012)

### ABSTRACT

In modern pedagogical processes various teaching methods and approaches (elements of the pedagogical process – EPPs) are used ranging from traditional ones (e.g., lectures, books) to more recent ones (e.g., e-discussion boards, e-quizzes). Different models for evaluation of the appropriateness of EPPs have been proposed in the past. However, the literature shows that these models typically focus only on the appropriateness of a single EPP and do not provide information about its relative appropriateness in relation to other EPPs. Unfortunately, this considerably limits the use of such evaluation models for the needs of the educational institutions' management. In order to decide which EPPs to promote or modify, management requires a comparative overview of the appropriateness of all EPPs that are part of the pedagogical process under consideration. Therefore the goal of our study was to design a model which would facilitate a comparative evaluation of many e-learning and traditional EPPs by simultaneously considering perspectives of students' and teachers' who participate in a certain pedagogical process. We applied the proposed model to three real-life pedagogical processes that are presented in this paper. Three groups of students, their teachers, and the college's management participated in the study. The management confirmed that the evaluation model provided them with valuable information in order to plan actions for improvement of the pedagogical processes.

### Keywords

E-learning, Teaching methods, Tools, Evaluation, Model, Pedagogical process

### Introduction

When planning curricula it is useful to have a variety of teaching and learning methods and approaches to meet the diverse learning needs of the target audience (Haw, & Keating, 2006). In this paper, the term Element of Pedagogical Process (EPP) is used for such methods and approaches. Many studies discuss the possibilities of use of different e-learning EPPs in curricula (e.g., Fernández Alemán, Carrillo de Gea, & Rodríguez Mondéjar, 2011; Forsberg, Georg, Ziegert, & Fors, 2011; Gaberson & Oermann, 2010; Pucer, 2011; Starčič, 2008). With the advent of Web 2.0, many new interactive technologies which can be used in pedagogical process became available (e.g., e-quizzes, e-forums). Existing studies (Brown, 2011; Gabriela, 2009; Gunnar, 2009) have encountered benefits and problems related to the use of Web 2.0 in higher education. Moreover, the introduction of new e-learning EPPs can be lengthy and often requires a lot of effort. Therefore it is important to realize that the success of e-learning EPPs largely depends on the context. For instance, Brown (2011) argues that blanket application of Web 2.0 is not appropriate and that e-learning is useful in promoting student-centred learning, but not at all times and in all contexts. To understand the value of e-learning EPPs, it is important to evaluate them in relation to the actual context.

According to Haw and Keating (2006), the implementation of pedagogical processes and its EPPs should be appropriately evaluated. Different models and methods that measure the quality of curricula, quality of education, effectiveness of pedagogical processes, etc. already exist. Although these models can be used to help evaluate and improve the pedagogical process, they have two considerable limitations. First, they either evaluate the pedagogical process at a very high level (i.e. at the level of curricula or even institutions), omitting the details of individual EPPs (e.g. Haw, & Keating, 2006), or at a level of an individual EPP ignoring the pedagogical process as a whole (e.g., Fernández Alemán et al., 2011; Kelly, Lyng, McGrath, & Cannon, 2009). Richardson (2005) states there is no reason to think that obtaining feedback at one level would be effective in evaluating or improving quality at another level. Therefore it is important for teachers and management to obtain feedback at different levels simultaneously: at higher levels to understand the process as a whole and to compare different EPPs, and at lower levels to be able to understand single EPPs and propose tangible improvement actions. Second, many studies focus only on one group of

EPP users, typically teachers (e.g., Yuen & Ma, 2008) or students (e.g., Richardson, 2005) and study their attitudes towards a certain EPP. However, evaluation of an EPP from the students' or teachers' perspective alone is insufficient, because teachers might consider some EPPs as appropriate even if they are not well accepted by the students, while students might favour certain EPPs which are less effective or unpopular among teachers. As a result, improvement of the individual EPPs and the pedagogical process as a whole can be managed appropriately only through a combined understanding of students' and teachers' perspectives on e-learning and traditional EPPs and of the pedagogical process at different levels.

In order to improve the understanding of the adoption and value of EPPs used in the pedagogical processes we pose the following research questions: R1 - How can we model the evaluation of pedagogical process by concurrently considering students' and teachers' perspectives on e-learning and traditional EPPs? R2 - Can such model provide relevant information about e-learning and traditional EPPs previously unknown to teachers and management? R3 - Can such model be used to identify less effective or ineffective e-learning and traditional EPPs that should be improved or replaced? R4 - Can such model be used to compare the effectiveness of e-learning and traditional learning EPPs? R5 - Can such model help to improve the pedagogical process?

These questions are addressed in the continuation of the paper as follows. *Background* section presents a review of previous studies, followed by *Evaluation model* section, which presents our model in detail. The application of our model in three real pedagogical processes is presented in *Application of the evaluation model in practice* section, followed by sections *Discussion* and *Conclusion*.

## Background

In order to construct an appropriate evaluation model a thematic review of the literature was undertaken. Several studies evaluate the quality of the pedagogical process at the level of EPPs (Kelly et al., 2009) including studies which compare the effectiveness of a certain e-learning EPP to the effectiveness of traditional EPP (Fernández Alemán et al., 2011). Unfortunately, these studies typically evaluate only one EPP or at most compare two EPPs (Campbell, Gibson, Hall, Richards, & Callery, 2008), but do not compare the quality of many different EPPs used within a particular pedagogical process. Hence, it is difficult to objectively compare the effectiveness of different EPPs, which is fundamental for the improvement of the pedagogical process. A similar situation was detected in the field of software development process evaluation by Vavpotic and Bajec (2009), who also proposed a solution for this situation. Although pedagogical processes and software development processes are quite different in content, they share many similarities in terms of their general structure and they both can be considered as specialised types of business processes consisting of different process elements. For instance, in both process types we can find people who perform certain roles (e.g., teachers / programmers), activities that are performed by these roles (e.g., lecturing / programming), tools that help perform these activities (e.g., books / software modelling tools), etc. All these process parts can be evaluated and eventually changed to improve the process. Hence, we used the idea proposed by Vavpotic and Bajec (2009) that was already successful in the field of software development processes to evaluate the pedagogical process and its EPPs.

Currently, several models and methods can be found in the literature that assure the quality of pedagogical processes and provide documentation for evaluation and accreditation processes (Little, 2009). They often consider the following dimensions of pedagogical processes or their EPPs: opportunity for use, quality of knowledge gained, and student's level of acceptance. There have been a variety of studies focusing on the acceptance of e-learning by students (Yu, Chen, Yang, Wang, & Yen, 2007; Žvanut et al., 2011), discussing the opportunities of using e-learning in pedagogical processes (e.g., Emerson, 2007; Moule, Ward, & Lockyer, 2010; Ruiz, Mintzer, & Leipzig, 2006) comparing e-learning versus traditional learning on the acquisition and retention of knowledge (e.g., Campbell et al., 2008; Fernández Alemán et al., 2011) and evaluating the quality of knowledge gained by the students through the use of e-learning and the level of student's acceptance of e-learning (e.g., Kelly et al., 2009). However, all these studies consider only one or at most two dimensions, but do not consider all three dimensions concurrently.

The literature review shows that opportunities for use, quality of knowledge gained and student's level of acceptance are commonly evaluated for e-learning and traditional EPPs. However, none of the previous studies consider all of these dimensions concurrently, and they only focus on one or few specific EPPs. Based on these findings, we

propose an evaluation model that facilitates the concurrent evaluation of the three key dimensions for different EPPs used in the pedagogical processes.

## Evaluation model

### Structure of the evaluation model

To evaluate a pedagogical process we propose a model that comprises three main dimensions (Figure 1). The first dimension is the frequency of opportunities for use of an EPP. This dimension is evaluated by teachers who participate in the pedagogical process. It does not consider how often the EPP is actually used but only evaluates the perceived number of opportunities for the use of the EPP. EPPs with higher opportunities for use play a more important role in the current pedagogical process. However, that does not necessarily mean that EPPs with lower opportunities for use are not important. In some cases it is possible to alter the pedagogical process in order to increase the opportunities for use of the EPP that is otherwise evaluated as valuable. The dimension is grounded on existing studies as discussed in the Background section and is evaluated through a single property that is adopted from the model proposed by Vavpotic and Bajec (2009).

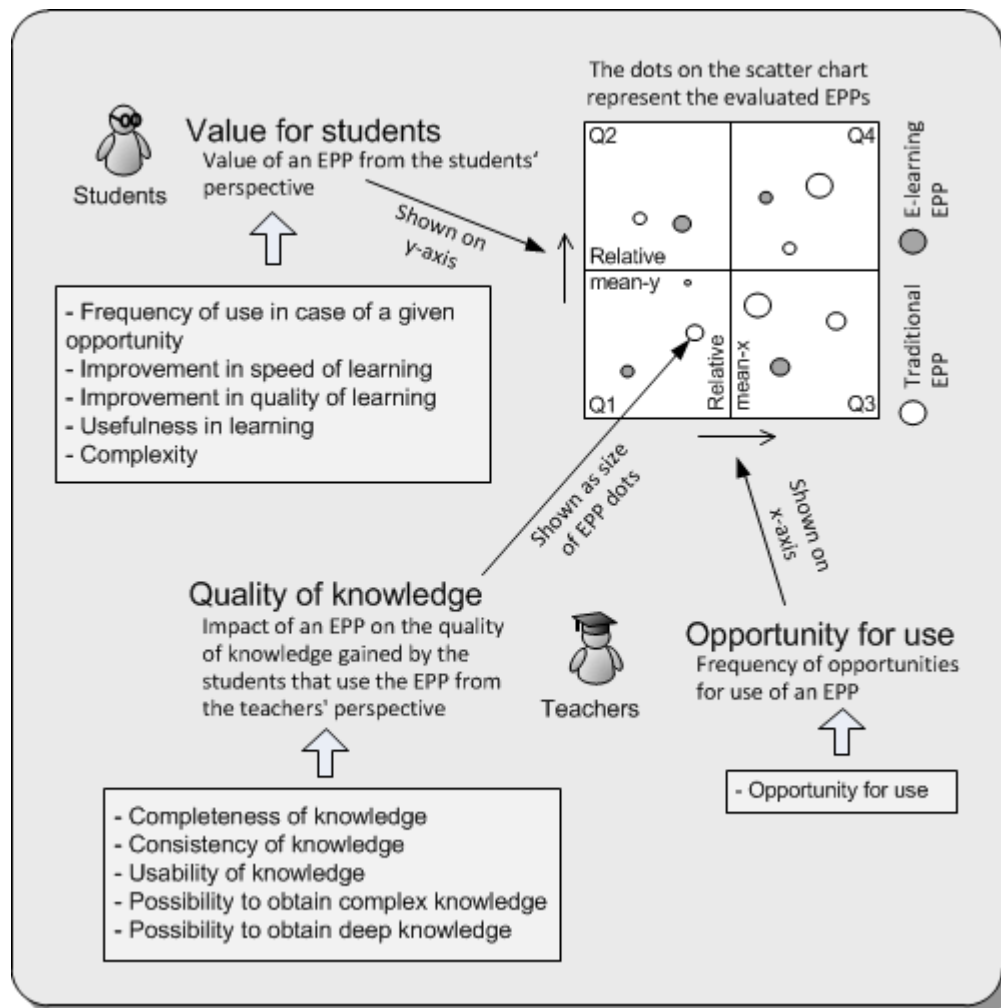


Figure 1. The evaluation model and presentation of the results

The second dimension is the impact of an EPP on the quality of knowledge gained by the students that use the EPP. A theoretically sound approach would be to create several individual groups of students (each group using the EPP that other groups do not use) and compare their exam results. Unfortunately, such an approach is not feasible in



practice as a large number of EPPs under evaluation would require the formation of many individual groups of students which would consequently be too small to produce conclusive results. Thus, this dimension focuses on the teachers' perceptions of the quality of knowledge gained rather than on actual quantitative data (e.g., exam results). Quality of knowledge is evaluated through five different properties, which consider both declarative and functional dimensions of knowledge (Biggs & Tang, 2011): the completeness, consistency and usability of obtained knowledge, and the possibility to obtain complex and deep knowledge.

The third dimension is the value of an EPP from the students' perspective. Five properties were used to measure this dimension. The first property is the EPP's frequency of use in case of a given opportunity. It is measured in two ways depending on whether the EPP under consideration is mandatory (e.g., mandatory lectures) or non-mandatory (e.g., participation in e-learning forums). In the case of non-mandatory EPPs, students report how often they actually use the EPP during their study, while in the case of mandatory EPPs they report how often they would use the EPP during their study even if it would not be mandatory. In this manner, the students' perceived acceptance of a particular EPP is measured regardless of whether the EPP is mandatory or not (adopted from Vavpotić & Bajec, 2009). Additionally, to better understand the reasons for the EPP acceptance level the students evaluate four properties grounded on the diffusion of innovations theory (DOI) (Rogers, 2003). Although DOI uses several different predictors of innovation diffusion, we only focus on two, namely relative advantage and complexity, as our preceding study (Žvanut et al., 2011) showed that only these two predictors have a significant impact on the nursing student's adoption of the EPPs. Relative advantage is evaluated through three properties, namely improvement in quality, improvement in speed, and usefulness, while complexity is measured through a single property. These properties are derived from existing studies which use similar properties to measure relative advantage and complexity (Riemenschneider, Hardgrave, & Davis, 2002; Vavpotić & Bajec, 2009; Vavpotić & Hovelja, 2012).

The concurrent evaluation of the three dimensions gives us a comprehensive understanding of the EPP's value from students' and teachers' perspectives. Such an understanding provides the basis for the improvements of the EPPs and the pedagogical process as a whole.

### **The measurement instrument**

The measurement instrument comprises two questionnaires (one for students and one for teachers). Both questionnaires use closed-ended questions where frequency of opportunities for use (teachers' questionnaire) and frequency of use in case of a given opportunity (students' questionnaire) were evaluated on a 7-point ordinal scale (never=1, very seldom=2, seldom=3, sometimes=4, often=5, very often=6, always=7) while all other items of both questionnaires used a 7-point Likert scale between 1 (strongly disagree) and 7 (strongly agree). For each EPP the teachers' questionnaire consists of 6 questions and the students' of 5, respectively (Figure 1).

To assure the content validity of the questionnaire, three experts in the field of pedagogy re-examined the questionnaire. The questionnaires were pilot tested on a group of 21 students and 4 teachers for 10 commonly used EPPs. This group also examined the items of both questionnaires for clarity. Both dimensions that comprise multiple properties of the quality of knowledge gained and value for students showed good internal consistency. The correlations between the properties of these two dimensions and the opportunity for use dimension were significantly lower than the internal correlations. These results were later reconfirmed on real life cases presented in the continuation of the paper.

### **The method of application of the evaluation model**

The method of application of the evaluation model is organized in three phases. The first phase is to catalogue the EPPs used in a pedagogical process. In order to identify them correctly, it is important that the teachers involved in the evaluated pedagogical process participate in focus groups, where EPPs are catalogued. In the second phase, teachers and students, the participants of the evaluated pedagogical process, evaluate the catalogued EPPs. In the third phase the results of the evaluation are individually analysed and discussed with the members of the college's management (i.e. the dean, the vice dean, the head of the nursing department, and the president of the governance board).

A scatter chart (Figure 1) shows the opportunity for use and value for students on horizontal and vertical axes, respectively. The evaluated EPPs are shown as numbered dots on the scatter chart, where the quality of knowledge gained is represented by the size of the dots. Relative means of opportunity for use and value for students divide the scatter chart into four quadrants (Figure 1: Q1 – Q4). Relative means are used since we are interested in EPPs' positions relative to other EPPs to be able to identify improvement opportunities. Q1 contains EPPs with low opportunity for use and low value for students, Q2 contains EPPs with high value for students but low opportunity for use, Q3 has EPPs with low value for students but high opportunity for use, and Q4 shows EPPs with high opportunity for use and high value for students.

In order to improve the pedagogical process teachers and management should focus primarily on the EPPs that have a considerable positive impact on the quality of knowledge and are positioned in either Q2 or Q3. For the EPPs in Q2 (e.g., Figure 2, EPP 10), teachers and management should investigate whether it is possible to alter the pedagogical process to create more opportunities for the use of such EPPs or to introduce new similar EPPs with more opportunities for use. In case of the EPPs in Q3 (e.g., Figure 2, EPP 8), further investigations are required and actions should be taken to improve their value for students by either properly presenting them to the students or by replacing them with EPPs that are more acceptable from the students' perspective. Different actions should be taken for the EPPs in Q1 (e.g., Figure 2, EPP 13). These EPPs neither have high value for students nor high opportunity for use. Consequentially, teachers should determine whether these EPPs can be replaced by other EPPs or removed from the pedagogical process. Finally, the EPPs in Q4 (e.g., Figure 2, EPP 5) that have high opportunity for use and value for students should be periodically monitored in order to identify significant changes of their position.

The positioning of EPPs on the scatter chart helps to identify the EPPs which should be improved and indicates the general cause of action for the improvement of EPPs. However, actual improvement scenarios are based on additional discussion and in-depth analysis of the use of EPPs. Examples of such analyses are presented in the following section.

## Application of the evaluation model in practice

### Overview

We used an embedded case study design as defined by Yin (2009) to evaluate the EPPs of the three nursing pedagogical processes that represented our units of analysis. The investigation took place at the College of Health Care Izola, University of Primorska, Slovenia at the end of the 2010/2011 academic year before the first exams. The college is located in a bilingual territory (Slovenian and Italian are the official languages), located 10 km from Italy. After the ethical commission of the college confirmed the research plan and questionnaire, the request for approval was sent to the management board, which approved and supported our study.

*Table 1. EPPs identified in the focus groups*

No. in scatter chart	EPP	EPP type
1	role play	traditional
2	e-learning (as an activity)	e-learning
3	laboratory practice (nursing)	traditional
4	traditional lectures	traditional
5	clinical practice	traditional
6	Consultations	traditional
7	laboratory practice (non-nursing)	traditional
8	Seminars	traditional
9	Books	traditional
10	simulation mannequin	traditional
11	educational films	traditional
12	e-mail	e-learning
13	e-quizzes	e-learning

14	e-discussion board	e-learning
15	case study	traditional
16	digital learning material	e-learning
17	team work	traditional

The three pedagogical processes were performed at two different units of the college, attended by three different groups of students, and supervised by common management. Our embedded units of analysis represented three commonplace situations in undergraduate education: a pedagogical process attended by full-time students, a pedagogical process attended by part-time students, and a pedagogical process attended by full-time students at the college's remote unit. Although all three pedagogical processes were based on the same curriculum, different teachers participated in each pedagogical process. Therefore for each pedagogical process a focus group of three teachers was formed, where two main groups of EPPs used in the pedagogical process were identified, namely activities and tools. These teachers also evaluated the quality of knowledge and the opportunity for use of EPPs that were part of the pedagogical processes in which they participated. Table 1 presents the identified EPPs with their respective numbers in the scatter charts (Figures 2 – 4).

In the following subsections we discuss the most relevant results of the three situations to demonstrate how the model provided relevant information about e-learning and traditional learning EPPs (R2), helped to identify less effective or ineffective e-learning and traditional learning EPPs (R3), was used to compare the effectiveness of e-learning and traditional learning EPPs (R4), and helped to improve the pedagogical process (R5). The evaluations of only selected representative EPPs are discussed, although other EPPs were examined in a similar manner.

### Situation 1

The first group consisted of first year full-time students, who attended the study at the remote unit, approximately 100 km from the college's headquarters. Even though the same undergraduate study programme of nursing took place in this unit as at the headquarters, the students in this group suffered from a substantial lack of resources required for the study (e.g., distance to the library, lack of support from teachers and administrative staff). The group consisted of 38 students from Slovenia and Italy: 10 males and 28 females. Their average age was  $20.0 \pm 1.2$ . The results are presented in Figure 2.

EPP 9 (books, positioned in Q3) was evaluated as having relatively high opportunity for use while value for students was below the relative mean. Discussion with the dean and the head of the nursing department showed that the distance from the library located at the college's headquarters presented a serious obstacle for this group of students. Consequentially the EPP 16 (digital learning materials) had higher value for students than EPP 9 as EPP 16 was used in many courses and it was easily accessible through the Internet (R2, R3, R4).

Surprisingly, EPP 2 (e-learning as an activity, positioned in Q1) had relatively low value for students and opportunity for use. This EPP plays an important role in this group as it bridges the spatial gap between the unit and the headquarters. The discussion with all the interviewed management confirmed the validity of our results. For instance, the vice dean, responsible for the quality of pedagogical processes, noted that the use of this EPP was discussed with teachers in this unit on several occasions. The teachers generally did not favour this EPP, as a lot of effort is required to prepare the e-contents and to promote its use. However, both the vice dean and the head of the nursing department were previously unaware of the EPP's low value for students (R2, R3). Both of them confirmed that this result would serve as an argument for justifying promotion activities about this EPP in the future (R5).

EPP 17 (team work, located in Q1) has value for students and opportunity for use below their relative means. Generally, teamwork and other social skills are very important in clinical practice. Hence, the opinion of the vice dean and the head of the nursing department was that this EPP should not be omitted from the pedagogical process. The discussion showed that they were previously unaware of this problem and that the use of this EPP would have to be substantially reconsidered (R2, R3). In order to better understand this situation, the dean suggested conducting an additional investigation among students and teachers (R5).

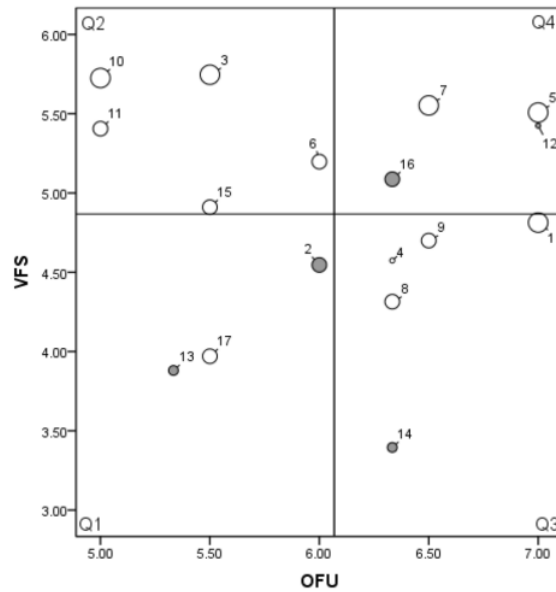


Figure 2. Situation 1 – results

## Situation 2

The second group consisted of first year regular students, who attended the pedagogical process at the college's headquarters, where a plethora of resources is available for study (e.g., full time presence of the entire staff for consultation and technical support, library, computer classroom, etc.). The group consisted of 48 students from Slovenia and Italy: 11 males and 37 females; their average age was  $19.8 \pm 0.8$ .

In this group, EPP 15 (case study) fell into Q1, as its value for students and its opportunity for use are both below their relative means (Figure 3). The vice dean noted that they were previously unaware of the low opportunity for use and low value for students of this EPP (R2). As in Situation 1, the vice dean confirmed that these results represent a valid argument for future activities to increase the opportunities for use and improve the value for students of this EPP (R5).

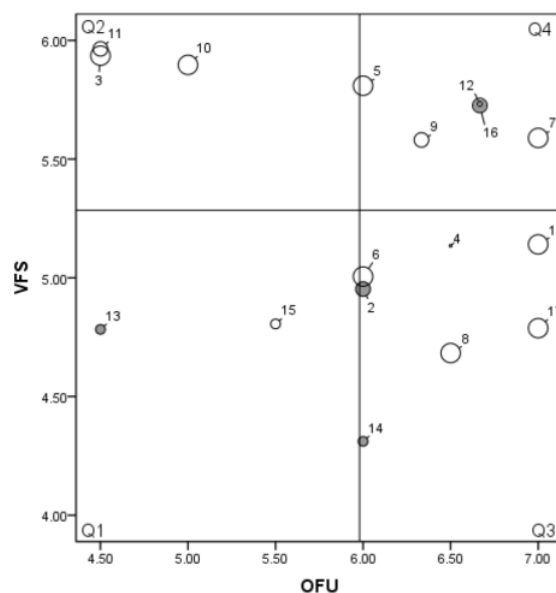


Figure 3. Situation 2 - results

EPP 2 (E-learning as an activity) fell in Q3 as the value for students is below and opportunity for use slightly above their relative means. A relatively low value for students was detected also for e-learning EPP 13 (e-quizzes) and EPP 14 (e-discussion boards). The dean and the vice dean again expressed surprise over the low value for students of these EPPs. According to their experience, first year students often complained about the EPP 5 (clinical practice) and EPP 3 (laboratory practice). However, the results of the evaluation revealed that the aforementioned traditional EPPs have considerably higher value for students than the three e-learning EPPs. This was contrary to the management belief that e-learning EPPs are generally very popular among students (R4). On the basis of these results, the vice dean, the head of the nursing department, and the president of the governance board suggested to conduct a further investigation on the value of e-learning EPPs in order to decide which of them should be promoted or omitted from the pedagogical process in the future (R5).

### Situation 3

The third group consisted of first year part time students. As in the preceding group, they attended their study at the college's headquarters, but only two days per week. The members of this group were older students who already had certain work experience in nursing. Their average age was  $31.1 \pm 6.5$ . This group consisted of 30 students from Slovenia: 3 males and 27 females. The critical problem of this group was lack of time, as they had to combine their work and study activities.

EPP 14 (e-discussion board) fell into Q3 and resulted in the EPP with the lowest value for students (Figure 4). This result came as a surprise to the vice dean (R2, R3). According to the vice dean's experience part-time students were usually more aware of the importance of social skills required for working in clinical practice than regular students. This observation was confirmed by evaluation of traditional EPPs like EPP 1 (role play) and EPP 17 (team work), which are typical EPPs that help the students to gain social skills. As expected, both EPPs fell into Q4, contrasting with the EPP 14 (e-discussion board). The vice dean and the head of the nursing department noted that the most probable reason for these results was that the this group of students perceived traditional EPPs as more appropriate for learning social skills than e-learning EPPs due to lack of direct contact (R4). To improve this situation the dean and vice dean decided to instruct teachers to better present the EPP *e-discussion board* to this group of students (R5).

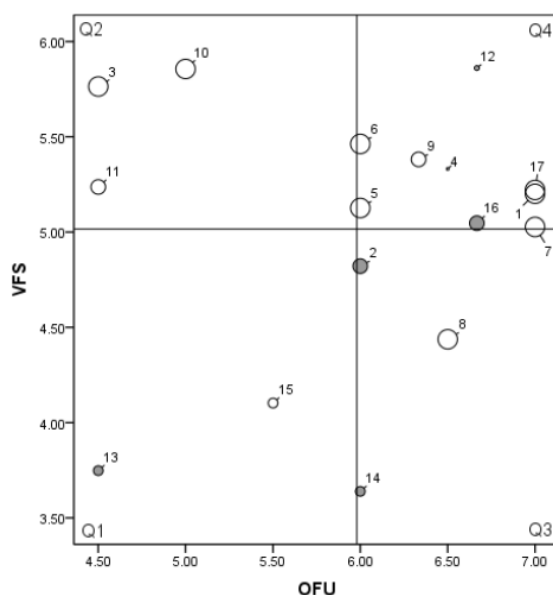


Figure 4. Situation 3 - results

EPP 3 (laboratory practice, positioned in Q2) had high value for students, relatively low opportunity for use and high quality of knowledge. The vice dean and the head of the nursing department noted that they were previously unaware of the high value for students of this EPP (R2). The vice dean suggested to increase the opportunity for use of this EPP as this group suffered from lack of time for studying. The vice dean's opinion was that the reason for the low

opportunity for use was in the relatively low number of hours devoted to this EPP in the curriculum. President of the governance board suggested that according to these results, the number of hours of laboratory practice in nursing should be reconsidered, as laboratory practice has a substantially lower cost than clinical practice in this institution. All members of school management concluded that this information represented a valid argument for a modification of the curricula for this group of students (R5).

## Discussion

The evaluation model was applied in three typical situations, where it was successfully used to identify less effective EPPs (R3) and to compare e-learning and traditional EPPs (R4). The results of the evaluations were carefully analysed and discussed with the members of the college's management, who confirmed that the evaluation model helped them to considerably improve their understanding of the three pedagogical processes (R2) and consequentially reach better decisions related to their improvement (R5). For instance, all higher education institutions in the EU have to periodically reaccredit their curricula. Reaccreditation provides an important opportunity for its modification. However, according to the college's management, decisions about the EPPs used in the curricula are often subjective. Therefore, the presented model can be used as a valuable tool that helps to bridge this gap by identifying less effective e-learning and traditional EPPs (R3), thus providing a more objective basis for the modification of the curricula (R5).

The concepts and processes of continuous quality improvements (CQI) are becoming relevant for the higher education (Heydman, 2006). Biggs and Tang (2011) suggest that in order to improve the pedagogical process, the focus should be shifted from teachers to students (i.e. student-centred teaching). Hypothetically, when students' wishes and opinions are fully considered there is a latent danger of excluding teachers' opinion about the use of a particular EPPs. However, excluding students' opinion could lead to instructor-centred teaching, where the teacher is the expert on the content and the delivery approach (Young & Maxwell, 2007). This lowers the flexibility of the pedagogical process, which is very important in a rapidly changing environment. The presented model considers both students' and teachers' opinions and helps the institution management and teachers to search for compromises, when decisions about the use of an EPP are under consideration. This is in accordance with CQI, where it is recommended that cross-functional teams assess whether the systems are optimal to produce best practices and results (Heydman, 2006). Our model, if used continuously, could represent a contribution to the CQI of an institution.

The importance of the presented model was confirmed through its application in practice which showed that several initial expectations of the college's management were misaligned with actual students' evaluations. For instance, some e-learning EPPs were quite unpopular among students even at the remote organizational unit where access to certain traditional EPPs was severely limited (R4). Although the management expected that e-learning EPPs would serve as a substitute for the less accessible traditional EPPs, the results of the evaluation showed that the students did not perceive some of the e-learning EPPs as valuable. The model enabled management to detect this situation and to use its results in further improvements of the pedagogical processes (R5). Similar misalignment was detected also in the teachers' and students' perceptions of certain EPPs. On one hand, the teachers perceived these EPPs as having significant positive advantages for students, but on the other hand, the students' level of adoption was relatively low. These findings are consistent with the DOI theory which states that even when an innovation has obvious advantages it is not always diffused and adopted rapidly (Rogers, 2003).

In practice, students typically have the opportunity to evaluate the pedagogical process at the course level and only rarely at the level of EPPs. Although evaluation at the level of the course suffices for the purpose of general course benchmarking, it does not provide detailed information about particular EPPs and their effectiveness in different courses. Consequently, the selection of suitable EPPs is typically left to the judgement of teachers and management.

An important advantage of the evaluation at the level of EPPs is that it facilitates application of the DOI theory to understand the students' personal adoption of different EPPs. A whole course or even a whole pedagogical process is typically too broad and complex to be viewed as a single innovation. Furthermore, the students' perceptions of the course or the pedagogical process are typically very limited and are often affected by their recent positive or negative experience with a certain part of the course or the pedagogical process. This severely inhibits the possibility for the application of the DOI theory on the level of the course or the pedagogical process as the students are not able to objectively evaluate them. However, our study demonstrates that students perceive a properly defined EPP as a

single innovation which they are able to objectively evaluate. With the development of information and communication technologies, innovations are continuously being introduced into pedagogical processes. In our opinion, the use of our model could be easily extended to additional pedagogical process innovations under the condition that these innovations could be defined at the level of the EPPs, i.e. at the appropriate level of granularity.

When using the presented model, the institution should take into consideration several limitations. Bastable (2008) presents three groups of barriers to conducting evaluations: lack of clarity, lack of ability, and fear of punishment and loss of self-esteem. In order to avoid or minimize the lack of clarity, a clearly stated purpose must explain why the evaluation is being conducted and what is being evaluated to all evaluation participants (i.e., the students, the teachers, and the management). All the evaluation participants should be motivated to participate and should recognize that their evaluation is essential for the improvement of the pedagogical process. The EPPs under evaluation must be properly presented to the students so they can fully understand and objectively evaluate them. To minimize the negative effects of lack of ability, the evaluators should be familiar with the presented model and have moderating skills and experience in order to guide the teacher focus groups who identify the EPPs used in the pedagogical process. Fear of punishment and loss of self-esteem could represent a problem for teachers participating in focus groups, as their participation is not anonymous as in the case of students. The evaluator of the pedagogical process has to take this into account when guiding the focus groups, as some teachers (e.g., newly employed teachers, old-fashioned teachers) are often not prepared to discuss the use of EPPs with other colleagues. Lee, Cerreto, and Lee (2010) in their study report that the subjective norm is a determinant of teachers' intentions to use the technology. It is not a rare situation that opinion leaders in some old-fashioned groups of teachers inhibit the innovative colleagues to express their opinion and present the innovative EPPs they used in practice. However, a skilled and independent evaluator can successfully manage such a situation by combining focus groups with individual informal interviews. Finally, the unit of analysis and the evaluation participants should be carefully chosen (Yin, 2009). In order to get reliable information from the evaluation, it is important that both students and teachers evaluate only the pedagogical process that they are actually involved in.

## Conclusions

In this paper, we proposed a model for comparative evaluation of e-learning and traditional EPPs. In comparison to the existing evaluation models our model has two distinguishing capabilities: it concurrently considers students' and teachers' perspectives and it facilitates the comparative evaluation of different EPPs that comprise the pedagogical process. These permit the users of the model to observe the value of pedagogical processes in detail, detect less suitable e-learning and traditional EPPs, and formulate focused actions for the improvement of pedagogical processes through improvement of their EPPs. The model was used to evaluate three commonplace pedagogical processes in the field of nursing, where it provided valuable information to the college's management who used it to direct their improvement efforts.

Our future work will focus on the application of the proposed model in a longitudinal manner in order to continuously monitor and improve the pedagogical processes. We will focus on questions such as: how frequently should the re-evaluation of a pedagogical process be performed, how different generations of students evaluate different EPPs, and how the context of use affects the dimensions used in our model, especially when e-learning EPPs are under consideration.

## Acknowledgements

The authors wish to thank all the students who participated in this study, the management of College of Health Care Izola, University of Primorska, Slovenia, which approved and supported this study, as well as Dr. Cyprian Laskowski, for careful reading and useful suggestions.

## References

Bastable, S. B. (2008). *Nurse as educator: principles of teaching and learning for nursing practice*. Boston, MA: Jones and Bartlett.

- Biggs, J. B., & Tang, C. (2011). *Teaching for quality learning at university: What the student does* (4th ed.). Maidenhead, England: Open University Press
- Brown, S. A. (2011). Seeing Web 2.0 in context: A study of academic perceptions. *The Internet and Higher Education*, 15(1), 50–57.
- Campbell, M., Gibson, W., Hall, A., Richards, D., & Callery, P. (2008). Online vs. face-to-face discussion in a web-based research methods course for postgraduate nursing students: A quasi-experimental study. *International Journal of Nursing Studies*, 45(5), 750-759.
- Emerson, R. J. (2007). *Nursing education in the clinical setting*. St. Louis, MO: Mosby/Elsevier.
- Fernández Alemán, J. L., Carrillo de Gea, J. M., & Rodríguez Mondéjar, J. J. (2011). Effects of competitive computer-assisted learning versus conventional teaching methods on the acquisition and retention of knowledge in medical surgical nursing students. *Nurse Education Today*, 31(8), 866–871.
- Forsberg, E., Georg, C., Ziegert, K., & Fors, U. (2011). Virtual patients for assessment of clinical reasoning in nursing - A pilot study. *Nurse Education Today*, 31(8), 757-762.
- Gaberson, K., & Oermann, M. (2010). *Clinical teaching strategies in nursing* (3rd ed.). New York, NY: Springer Publishing Company.
- Gabriela, G. (2009). To use or not to use web 2.0 in higher education? *Procedia - Social and Behavioral Sciences*, 1(1), 478-482.
- Gunnar, A. (2009). Web 2.0, pedagogical support for reflexive and emotional social interaction among Swedish students. *The Internet and Higher Education*, 13(4), 197-205.
- Heydman, A. M. (2006). Planning for accreditation: Evaluating the curriculum. In S. B. Keating (Ed.), *Curriculum Development and Evaluation in Nursing* (pp. 297-315). Baltimore, PA: Lippincott Williams & Wilkins.
- Haw, M. A., & Keating, S. B. (2006). Adaptation of curriculum development and evaluation of staff development and client-centered health education. In S. B. Keating (Ed.), *Curriculum Development and Evaluation in Nursing* (pp. 319-348). New York, NY: Lippincott Williams & Wilkins.
- Kelly, M., Lyng, C., McGrath, M., & Cannon, G. (2009). A multi-method study to determine the effectiveness of, and student attitudes to, online instructional videos for teaching clinical nursing skills. *Nurse Education Today*, 29(3), 292-300.
- Lee, J., Cerreto, F. A., & Lee, J. (2010). Theory of planned behavior and teachers' decisions regarding use of educational technology. *Educational Technology & Society*, 13(1), 152-164.
- Little, B. B. (2009). Quality assurance for online nursing courses. *Journal of Nursing Education*, 48(7), 381–387.
- Moule, P., Ward, R., & Lockyer, L. (2010). Nursing and healthcare students' experiences and use of e-learning in higher education. *Journal of Advanced Nursing*, 66(12), 2785-2795.
- Pucer, P. (2011). Augmented reality. *Slovenian Medical Journal*, 80(7-8), 578-585.
- Richardson, J. T. E. (2005). Instruments for obtaining student feedback: a review of the literature. *Assessment & Evaluation in Higher Education*, 30(4), 387-415.
- Riemenschneider, C. K., Hardgrave, B. C., & Davis, F. D. (2002). Explaining Software Developer Acceptance of Methodologies: A Comparison of Five Theoretical Models. *IEEE transactions on software engineering*, 28(12), 1135-1145.
- Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York, NY: Free Press.
- Ruiz, J. G., Mintzer, M. J., & Leipzig, R. M. (2006). The impact of e-learning in medical education. *Academic Medicine*, 81(3), 207-212.
- Starčič, A. I. (2008). Developing virtual simulation game for authentic learning : Realizing partnership between university and industry. *WSEAS transactions on communications*, 7(7), 786-795.
- Vavpotic, D., & Bajec, M. (2009). An approach for concurrent evaluation of technical and social aspects of software development methodologies. *Information and Software Technology*, 51(2), 528-545.
- Vavpotič, D., & Hovelja, T. (2012). Improving the evaluation of software development methodology adoption and its impact on enterprise performance. *Computer Science and Information Systems*, 9(1), 165-187.
- Yin, R. K. (2009). *Case study research: design and methods* (4th ed.). Thousand Oaks, CA: Sage Publications.



Young, L. E., & Maxwell, B. (2007). Student-centered teaching in nursing: From rote to active learning. In S. Yu, I. J. Chen, K. F. Yang, T. F. Wang, & L. L. Yen (2007). A feasibility study on the adoption of e-learning for public health nurse continuing education in Taiwan. *Nurse Education Today*, 27(7), 755-761.

Yuen, A. H. K., & Ma, W. W. K. (2008). Exploring teacher acceptance of e-learning technology. *Asia-Pacific Journal of Teacher Education*, 36(3), 229-243.

Žvanut, B., Pucer, P., Ličen, S., Trobec, I., Plazar, N., & Vavpotič, D. (2011). The effect of voluntariness on the acceptance of e-learning by nursing students. *Nurse Education Today*, 31(4), 350-355.

## A Novel Method for Learner Assessment Based on Learner Annotations

**Fakhroddin Noorbehbahani\*, Elaheh Biglar Beigi Samani and Hossein Hadian Jazi**

Isfahan University of Technology, Department of Electrical and Computer Engineering, Isfahan, Iran //  
f.noorbehbahani@ec.iut.ac.ir // biglarbeigi@nsec.ir // hadian@nsec.ir

\*Corresponding author

(Submitted February 19, 2012; Revised July 22, 2012; Accepted August 27, 2012)

### ABSTRACT

Assessment is one of the most essential parts of any instructive learning process which aims to evaluate a learner's knowledge about learning concepts. In this work, a new method for learner assessment based on learner annotations is presented. The proposed method exploits the M-BLEU algorithm to find the most similar reference annotations and then the learner annotation will be processed further to check essential words, words order and contradictions. To examine this new approach, a virtual learning environment was designed and implemented in which assessment of the learner's knowledge is performed on the basis of main and sub concepts. These concepts are delivered by means of course contents and the learning environment guides the user to annotate concepts. Evaluation results show that our designed system can effectively assess learner's knowledge. The benefit of suggested assessment method is its implicitness of assessment approach. Furthermore the correct annotations can be used to help the users remembering concepts by reviewing their annotations.

### Keywords

Computer assisted assessment, Learner assessment, E-learning, Learner annotation

### Introduction

In light of the rapid development of distance learning, assessment functionality has become one of the most important key factors in online education (Belcadhi, Henze, & Braham, 2004). Presently, in most existing Learning Management Systems (LMS), assessment strategies are focused on simple question-and-answer methodologies. For example, single and multiple-choice questions (MCQ) with several answer choices, and radio-buttons for selecting the correct answers are the most popular ones (Belcadhi et al., 2004). On the other hand, annotations are a natural way for recording comments, ideas and summarizations in specific contexts within a document. When people review a paper or read a book, they often underline important parts of the document or write notes in the margin (Brush, 2002). Annotations being made by a learner in an online learning environment highlight important pieces of the content that can be accessed by other learners. This functionality of annotation is increasingly appreciated as an interesting tool in Web-Based Collaborative Applications (Petkovic et al., 2005). In the field of e-learning, making annotations is a natural way for recording and summarizing important contents in specific contexts within a document that effectively improves and speeds up the learning process.

Learners can study learning contents and annotate them in their own words and in the next time learner annotations help him/her to remember the concepts. In this research, for the first time, learners' annotations are utilized for learning progress assessment.

In this work, a novel assessment method is presented. This method is performed by applying a modified version of the BLEU algorithm on learner's annotations. The BLEU algorithm has been initially applied in evaluating machine translation systems (Papineni, Roukos, Ward, & Zhu, 2002). M-BLEU is a modified version of BLEU proposed by Noorbehbahani & Kardan, 2010. M-BLEU is valuable to score sentences and free text responses, but it does not consider contrast and other issues. We apply the M-BLEU algorithm to discover the most similar reference annotation to a learner annotation. Then the discovered reference annotation is compared with the learner annotation in a 3-step process for checking essential words, words order, and any contradictions. If for any specific learning concept, learner annotation passes all 3 steps, it shows the complete understanding of the learner about the concept being assessed and the score is 1, otherwise the score is 0.

The arrangement of the paper is as follows: in the related works section, some related works in the domain of Computer Assisted Assessment will be described. The next section is dedicated to the prerequisites necessary for annotation-based assessment. The prerequisites are covered under the Reference generation, Learner annotation

submission and Right side annotation paraphrases. Following those sections, the reference annotation selection procedure which uses M-BLEU will be discussed. The Annotation checking section is about refinement of assessment. To refine annotation-based learner assessment, a final examination should be done. This examination should be carried out in three steps, namely: checking essential words, checking words order, and checking contradictions. The evaluation of obtained results from the proposed assessment method will be discussed in the evaluation section. Finally in the last section, conclusion to this work will be presented.

## **Related work**

Assessment is an important stage in any educational process for evaluating the learner's knowledge regarding the concepts being learned. Hence, a new research branch called Computer Assisted Assessment (CAA) has been developed to study how computers can be utilized in the evaluation of learners' learning progress. The work being done by Essay Grader (PEG) is the first CAA system for assessing free text answers (Page, 1996). The PEG was intended to enhance the assessment process. In this system, an essay is predominantly graded by PEG based on writing quality, without taking content into account. The PEG is based on the concept of "proxes", i.e., computer approximations or measures of intrinsic variables of interest within the essay to simulate grading by a human rater. PEG applies a statistical approach based on the assumption that the quality of essays is reflected by the measurable proxes. In PEG the Natural Language Processing (NLP) technique is not used and lexical content is not considered (Putri Ratna, 2006).

The idea of utilizing computer for evaluating open ended questions was expressed by Webster (Webster, 1990). Webster applied more appropriate techniques rather than only using style checkers in his work.

In 1990, Educational Testing Service I was developed at the Educational Testing Service (ETS) organization in USA. It relied on more direct measures of writing quality and did not consider the content. ETS I only worked on short sentences (not longer than 20 words) (Whittington & Hunt, 1999). The Intelligent Essay Assessor (IEA) was developed at the University of Colorado in the USA. IEA focuses on the content, and it is based on Latent Semantic Analysis (LSA) (Foltz, Laham, & Landauer, 1999). In LSA, the words order is not considered because the authors claim that this is not the most important factor for acquiring the sense of a passage. LSA also requires large amounts of data to construct an appropriate matrix representation of words being used. Due to the size of the matrices, LSA requires a heavy computation (Putri Ratna, 2006).

E-rater is an enhanced version of Educational Testing Service I. In E-rater, a hybrid approach by combining NLP and statistical techniques is utilized. E-rater measures the organization, the sentence structure and the content (Burstein, Kukich et al., 1998).

IntelliMetric was created by the American company Vantage Learning Technologies. It applied an Artificial Intelligence approach for assessment of both the style and the content (Vantage Learning Tech., 2000). C-rater is a NLP based prototype that was developed for assessing short answers related to the content-based questions such as those may appear in a review section of a book chapter (Burstein, Leacock, & Swartz, 2001). Many of the NLP tools and techniques that had been developed for E-rater were adopted by C-rater. Rather than a large collection of graded answers for training, C-rater uses single correct answers.

In 1998 Larkey introduced her system based on text categorization techniques (TCT), text complexity features, and linear regression methods (Larkey, 1998). In 1999 Christie presented SEAR (Schema Extract Analyze and Report) which is based on pattern-matching techniques to provide automated marking of the essay content (Christie, 1999).

Another system is Automark which employs NLP techniques to perform an intelligent search among free text responses according to predefined computerized mark scheme answers (Mitchell, Russell, Broomhead, & Aldridge, 2002). The Intelligent Essay Marking System (IEMS) was presented in 2000. IEMS is based on the Pattern Indexing Neural Network. IEMS can be used both as an assessment tool, a diagnostic tool or a tutoring tool in many content-based subjects (Ming, Mikhailov, & Kuan, 2000). In 2001, research at University of Portsmouth (UK) led to the Automated Text Marker (ATM). ATM looks for concepts in the text and dependencies between them to score items independently; first the style and second the content (Callear, Jerrams-Smith, & Soh, 2001). The work being done by Rudner and Liang resulted in the Bayesian Essay Test Scoring sYstem (BETSY) in 2002. This system is based on

statistical analyses. BETSY incorporates the best features of PEG, LSA and E-rater. It also has several crucial advantages of its own (Rudner & Liang, 2002). In 2002, the Paperless School free text Marking Engine (PS-ME) was developed. PS-ME was designed to integrate with a Web-based Learning Management System (Mason & Grove-Stephenson, 2002). Because of processing requirements, essays cannot be graded by PS-ME in real-time. In 2003, two new systems were developed: Auto-marking and CarmelTC. Auto-marking is based on NLP techniques and pattern matching (Sukkarieh, Pulmand, & Raikes, 2003), and CarmelTC makes assessments of learners' writing with machine learning classification methods and a naive Bayesian classification (Ros'e, Roque, & VanLehn, 2003).

## Prerequisites for assessment

### Reference generation

To assess a learner's annotation, it is necessary to compare it with a reference annotation. Each concept in the domain model has several questions for assessment. These questions are created by the content author or the course instructor. One learning concept can be delivered in a sentence, a combination of sentences, a paragraph, or a combination of paragraphs. Therefore the related questions to assess this concept can be also in different levels. In Figure 1, different combinations of concepts and related questions have been illustrated.

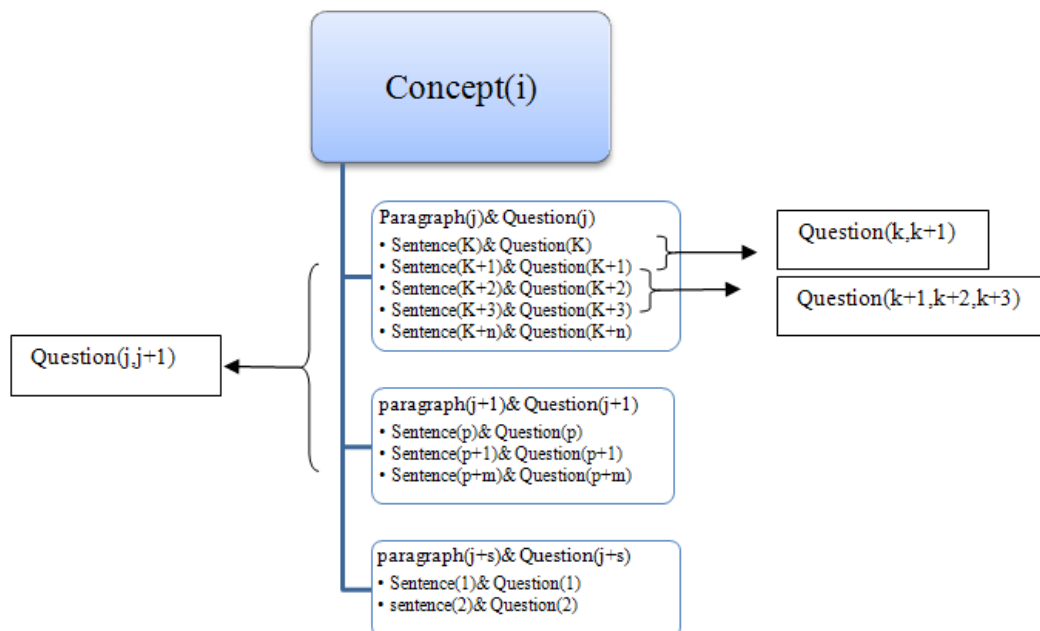


Figure1. Different combinations of concepts and related questions

For example, Sentence(K) has a question for assessment of its concept, namely Question(K). Sentence(K) to Sentence(K+3) have a combined Question(K+1,K+2,K+3) and Question(j) belongs to Paragraph(j).

Each Question and a related answer should be converted to a reference annotation. Every reference annotation in its complete format has two sides: a left side and a right side. Each question can be supposed as the left side of an annotation, and its related answer can be supposed as the right side of the same annotation. Some examples for converting questions and answers to reference annotations are shown in (Table 1):

As could be seen in Table 1, the reference annotation has a left and a right side. When a learner wishes to annotate about a specific concept, s/he clicks on the associated icon. Then system automatically generates the left side of the annotation and the learner should complete it by writing the right side. In the next step the learner annotation will be compared to the reference annotation.

Table1. Converting questions and answers to reference annotations Examples

Concept Question	Concept Answer	Generated reference annotation
What are the characteristics of online education?	Teacher and learner separation, Influence of an educational company, Use of computer network for educational content, Provision of 2 way communication	<u>Online education characteristics:</u> <ul style="list-style-type: none"> <li>Teacher and learner separation,</li> <li>Influence of an educational company,</li> <li>Use of computer network for educational content,</li> <li>Provision of 2 way communication</li> </ul>
Name 2 models for online education support systems.	Jigsaw model Hub model	<u>2 models for online education support systems:</u> <ul style="list-style-type: none"> <li>Jigsaw model</li> <li>Hub model</li> </ul>
What is an authoring tool?	Software application being used by non-programmers, that utilizes a metaphor (book, or flow chart) to create online courses	<u>Definition of authoring tool:</u> Software application being used by non-programmers, that utilizes a metaphor (book, or flow chart) to create online courses

### Learner annotation submission

Usually when study of part of the lessons is completed, the learner may want to annotate and summarize some of the learnt concepts. In such case, the learning concepts hierarchical tree is presented to the learner. When the learner clicks on a concept in the tree, the system automatically generates the left side of annotation. This action can help the learner to remember what has been studied by him/her earlier. Obviously the learner should complete the right side of annotation and submit it to the system. See the example in Figure 2.

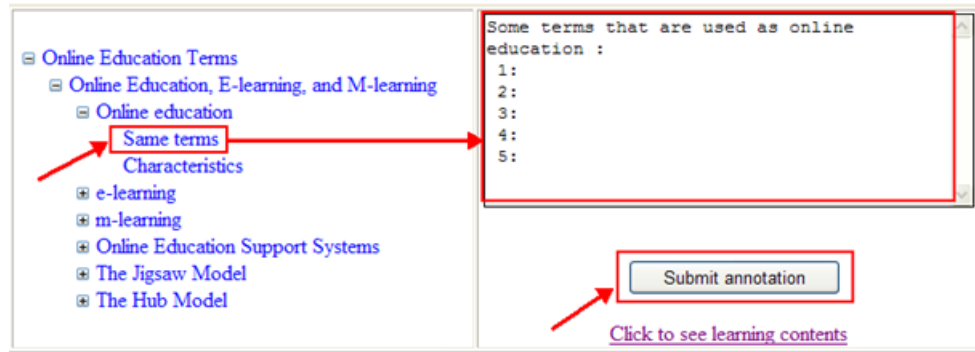


Figure2. Concept selection and annotation example

Each annotation may include several parts and each part covers a specific fraction of total knowledge. Considering the learner annotation as  $L = \{l_1, l_2, \dots, l_m\}$  and the reference annotation as  $R = \{r_1, r_2, \dots, r_m\}$ , each of  $L$  and  $R$  sets has  $m$  parts. The learner is guided by the e-learning system to complete  $m$  parts of an annotation. As shown in Figure 2, this sample annotation has 5 parts, and each part covers a fraction of the total knowledge. Let us assume that the fraction of the total knowledge being covered by each part is  $k_j$  and  $j = 1, 2, \dots, m$ . Each  $k_j$  is assigned with a number that represents the percentage (fraction) of knowledge. Therefore in the example illustrated in Figure 2, the annotation has 5 parts and all parts have the same percentage of knowledge (20% of total knowledge). When any learner successfully completes 2 parts of the mentioned annotation, s/he acquires 40% of the total knowledge related to a

specific concept. It is notable that if annotation has several parts, for each part several reference annotations are generated, and each part could be assessed.

### Right side annotation paraphrases

As mentioned, learner annotation should be compared with reference annotations. If there is only one reference annotation for each concept, the annotation assessment will fail. That is because a learner may annotate concepts in different syntaxes, keywords and word counts, while the meaning could be the same as the reference annotation. To increase the accuracy of the assessment, for each concept several reference annotations should exist. These reference annotations should differ from each other in syntax, keywords, and word counts, but they should have the same meaning. In other words, paraphrases should be generated as the reference annotations. Paraphrases are sentences that are equivalent in meaning with respect to the information conveyed by the content (Otterbacher & Radev, 2004).

In this research, paraphrases are generated by the experts. In this paper, our focus is on annotation assessment and therefore manually generating the reference annotations could result in accurate reference annotations. By preparing several reference annotations, the learner annotation can be compared to the most similar reference annotation. This results in a more accurate assessment.

### Reference annotation selection

To select the most similar reference annotation, we used M- BLEU method proposed by (Noorbehbahani & Kardan, 2010) which improves BLEU algorithm in 4 phases to make it useful to score free text answers.

For a reference annotation we should find a paraphrase of the reference annotation which is more similar in wording to the learner annotation than the original reference annotation. Consider the reference annotation as  $R$  and the learner annotation as  $L$ . A reference annotation  $S_{RL}$  that keeps the meaning of  $R$  having maximal word overlap with  $L$  should be generated by substitution of words from  $R$  with contextually equivalent words from  $S$ .

This substitution is based on two assumptions presented below:

- Words which exist in both the reference annotation and learner annotation are not considered for substitution. In fact, only unmatched words of the form  $\{r|r \in R - L\}$  and  $\{l|l \in L - R\}$  are taken into account.
- If  $n\text{-gram}_L \in L$ ,  $n'\text{-gram}_R \in R$  and  $n'\text{-gram}_R$  is the synonym of  $n\text{-gram}_L$ , then  $n\text{-gram}_L \in L$  will be substituted by  $n'\text{-gram}_R$  and  $n\text{-gram}_L$  is excluded from the next replacements. first  $3\text{-gram}_L$ , then  $2\text{-gram}_L$  and finally  $1\text{-gram}_L$  are considered for finding synonyms.

To increase performance, we use a domain dictionary for checking synonyms since there are unique technical vocabularies in each learning area. This domain dictionary, which should be updated and enhanced frequently, is generated by domain experts. Synonyms being used by top learners can also be used for improving the domain dictionary coverage. Since online education is the subject of the e-course selected for assessment of the proposed method, WordNet (Miller et al., 2006) was used as a general dictionary and E-learning glossaries are included in the domain dictionary.

To calculate similarity score, M-BLEU uses weighted n-gram precision calculated using Equation (1).

$$WP_{ra}(n) = \frac{\sum_{n\text{-gram} \in sa} \text{weight}_{ra}(n\text{-gram}) \times \text{Count}_{clip}(n\text{-gram})}{4 \times \sum_{n\text{-gram}' \in sa} \text{Count}(n\text{-gram}')} \quad (1)$$

Note that  $WP_{ra}^{(n)}$  is the weighted n-gram precision of the reference annotation and learner annotation  $weight_{ra}(n-gram) \in \{1, 2, 3, 4\}$ . The function  $Count_{clip}(n-gram)$  counts the number of n-gram clippings (or matches) between the learner annotation and the reference annotation. Finally,  $Count(n-gram)$  is the total number of n-grams in the learner annotation.  $Count_{clip}(n-gram)$  is calculated using Equation (2).

$$Count_{clip}(n-gram) = \min(Count_{la}(n-gram), Count_{ra}(n-gram)) \quad (2)$$

Note that  $Count_{la}(n-gram)$  is the number of times that an n-gram occurs in the learner annotation and  $Count_{ra}(n-gram)$  is the number of times an n-gram occurs in the reference annotation.

The modified BLEU (M-BLEU) score for the reference annotation is calculated using Equation (3).

$$M-BLEU_{ra} = \exp\left[\sum_{i=1}^N w_n \log(WP_{ra}(n))\right] \quad (3)$$

For calculating basic M-BLEU<sub>ra</sub> it is assumed that  $N = 4$  and  $w_n = 1/N$ .

To find the most similar reference annotation, a learner annotation should be compared with each reference annotation and the M-BLEU score should be calculated for each reference annotation independently. The reference annotation with the maximum score is then chosen to calculate the similarity score.  $Max_{Score}$  is calculated using Equation (4).

$$Max_{Score} = \max\{M-BLEU_{ref} \mid ref \in RA\} \quad (4)$$

Note that  $Max_{Score}$  is the maximum M-BLEU score and  $ref$  denotes the reference annotation.  $RA$  indicates the reference annotation set and  $M-BLEU_{ref}$  is the M-BLEU score for reference annotation  $ref$ . Therefore, Equation (5) describes  $Ref_{max}$ , which is a set of reference annotations that their differences with the maximum M-BLEU score is less than  $\theta$ .  $\theta$  is the threshold used to determine whether  $Max_{score}$  and  $M-BLEU_{ref}$  are equal. In this study, we set  $\theta = 0.1$ .

$$Ref_{max} = \{ref \mid ref \in RA \wedge |M-BLEU_{ref} - Max_{score}| \leq \theta\} \quad (5)$$

To consider word importance, the M-BLEU calculation should be based on different n-gram weights as described in (Noorbehhahani & Kardan, 2010). Meanwhile, the brevity penalty (BP) factor is utilized to select the shortest reference annotation among references with identical M-BLEU scores.

Equation (6) calculates the maximum  $BP$  between the reference annotations with identical M-BLEU scores.

$$Max_{BP} = \max\{BP_{ref} \mid ref \in Ref_{max}\} \quad (6)$$

Note that  $Ref_{max}$  is the set of reference annotations that have the maximum M-BLEU score calculated using Equation (5).  $BP_{ref}$  is the brevity penalty factor for the reference annotation  $ref$ .  $BP_{ref}$  is calculated using Equation (7) by considering  $l$  as the length of the reference annotation and  $c$  as the length of the learner annotation.  $Max_{BP}$  is the maximum  $BP$  among reference annotations with the maximum M-BLEU score.

(7)

$$BP = \begin{cases} 1 & \text{if } c > r \\ e^{(1-r/c)} & \text{if } c \leq r \end{cases}$$

Finally, the selected reference annotation  $SRA$  is chosen by Equation (8).

(8)

$$SRA = \begin{cases} \{Ref_{max}\} & \text{if } |Ref_{max}| = 1 \\ \{ref \mid BP_{ref} = \text{Max}_{BP}, ref \in Ref_{max}\} & \text{if } |Ref_{max}| > 1 \end{cases}$$

(9)

$$Score(SRA) = M\text{-}BLEU_{SRA} \times BP_{SRA}$$

If  $Score(SRA)$  calculated by equation (9) is greater than  $\beta$ , then we consider this learner annotation for further processing.  $\beta$  is the minimum similarity threshold for the selected reference and learner annotation. This threshold condition avoids this method from considering a reference annotation with a small similarity. We use a value of  $\beta = 0.7$  which has been obtained experimentally.

### Annotation checking

In the previous section, how a reference annotation is selected in comparison with a learner annotation was shown. To assess the knowledge of a learner, the steps illustrated in (Figure 3) should be taken:

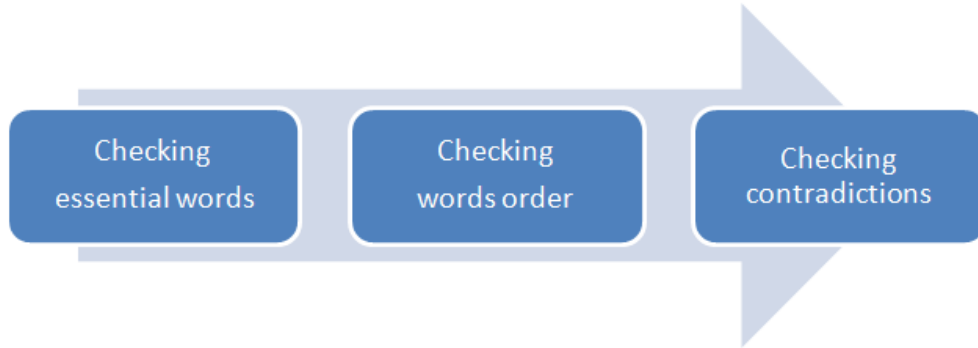


Figure3. The steps of final checking

If a learner annotation can pass all 3 steps, it means that the knowledge covered with the learning concept has been acquired by the learner. Each of the mentioned steps will be described fully in the next section.

### Checking essential words

In the previous section, selection process of a reference annotation with a maximum M-BLEU score in comparison with a learner annotation was shown. One problem with simply using the M-BLEU algorithm to evaluate learners' annotation is that it only considers the precision and ignores the recall; this means that it does not penalize those



learners' annotations that do not cover any fraction of the knowledge being covered by the reference annotation. To solve this problem, the existence of essential words of a reference annotation should be checked in a learner annotation.

In this paper, Essential words refer to the words of synthetic reference annotation ( $S_{RL}$ ), that if does not exist in the learner annotation; the learner annotation will be considered as incomplete. That means if one of the essential words does not appear in the learner annotation, it is probable that the learner does not have the knowledge of the learning concept. For example, consider this synthetic reference annotation:  
Definition of authoring tools: Any application software that allows individuals to create their own e-Learning content.

In this synthetic reference annotation, the words “application”, “create”, “e-Learning” and “content” are essential words. The Essential words of each reference annotation are specified by the content author or course instructor. When generating a synthetic reference annotation, if any essential word in the reference annotation could be replaced by its synonym, this synonym is also considered as an essential word.

If all essential words of a synthetic reference annotation are presented in the learner annotation, this annotation is passed to the next step.

### Checking words order

Sometimes changing the order of words can extremely change the meaning of a sentence. For example, consider the following sentences:

- Reference annotation: The Policeman killed the gunman.
- Learner annotation 1: The Policeman killed the gunman.
- Learner annotation 2: The gunman killed the policeman.

Due to the words order, the meaning of the sentence in learner annotation 1 has completely a different meaning from what is understood from learner annotation 2. Therefore, words order check is also a necessary procedure for learner annotation assessment. In our proposed method, annotations are divided into categories “L” and “H”. If an annotation belongs to category “L”, the order of its words has a Low importance and if it belongs to category “H”, the order of its words has a High importance for annotation-based assessment. When the instructor creates each of the reference annotation, s/he should categorize the reference annotation as category L or H. If an annotation belongs to “H” category, pairs of n-grams where order is important should be defined. A pair of  $(fromer_{n-gram}, latter_{n'-gram})$  indicates that  $fromer_{n-gram}$  should appear prior to the  $latter_{n'-gram}$  in the learner annotation.

It is notable that for words order similarity calculation, sometimes the n-grams' order is not considered. For example, time or place adverbs are not sufficiently important to be considered because changing their placements would not change the core meaning of a sentence.

If an annotation belongs to category “H”, the index of all n-grams pairs  $(fromer_{n-gram}, latter_{n'-gram})$  is checked.

### Checking contradictions

The last step to assess a learner annotation is checking contradictions. The Contradictions occur whenever the information being conveyed by two different texts is mismatched. The learner and reference annotations may have the same words order and syntax, but convey different meanings. For example, consider the following sentences:

- E-learning can improve education quality
- E-learning can never improve education quality.

In this example, two sentences are different in only one word, but they completely contradict each other.

There are 3 types of contradictions - negation, antonym, semantic and pragmatic information - that lead to contrast (Harabagiu, Hickl, & Lacatusu, 2006). Table 2 illustrates some examples of contradiction types. Noticeably, different types of contradictions can be seen between text A and text B.

Table2. Examples of different types of contradictions

Type of contradiction	Examples of contradiction
<b>Negation</b>	<b>Text A:</b> John held off a dramatic fight back from defending champion Andy. <b>Text B:</b> Defending champion Andy never took on John. <b>Comment:</b> Text A and B use predicates that have similar meanings, and employ the same arguments. But text B inverses the truth value through negation by using <i>never</i> .
<b>Antonym</b>	<b>Text A:</b> In California, 120 Central Americans began a hunger strike when their deportation was delayed. <b>Text B:</b> A hunger strike was called off. <b>Comment:</b> Contradiction caused by usage of the antonyms <i>begun</i> and <i>call off</i> .
<b>semantic and pragmatic information</b>	<b>Text A:</b> The explosion wounded the arm of Tom. <b>Text B:</b> Tom emerged unscathed from an explosion. <b>Comment:</b> Facts inferred from text A are denied by the facts inferred from text B that leads to contradiction

In the previous step, the existence of essential words of the reference annotation in the learner annotation is being checked. Obviously, any word in the reference annotation for which its antonyms could cause contradiction should also be considered as an essential word. If instead of essential words, their antonyms appear in the learner annotation, knowledge of learning concept is not considered for the learner due to the absence of essential words. The third type of contradiction RARELY takes place because the most similar reference annotation is selected for comparison with the learner annotation and then the essential words are checked. Hence, the annotation conveyed **semantic and pragmatic** information which cannot match with any reference annotation. Although the second and third types of contradiction occur rarely, they need a deep NLP for complete contradiction detection. Therefore, in this step only the contradiction of type 1 should be checked. Based on (Gindl, Kaiser, & Miksch, 2008), the negation classes being considered in this work are shown in Table 3.

Table3. Different types of negation classes

Type of negation	Example of negation	Negation keywords
<b>The Adverbial Negation</b>	<b>Comment:</b> This is the most frequent type of negation triggered by “not”, “n’t” and “never”. <b>Example:</b> Guideline developers do not recommend chemotherapy.	“not”, “n’t” “never”
<b>Intra-Phrase Triggered Negation</b>	<b>Comment:</b> In this negation type, “No” and “without” act as triggers. These triggers are included in the noun phrase. <b>Example:</b> Evidence obtained from at least one well-designed study without randomization	“No” “without”
<b>Prepositional Negation</b>	<b>Comment:</b> Triggers followed by prepositional phrases belong to this type of negation. This negation type often introduced by the prepositions “of” or “from”. The phrases following the preposition are considered as negated. Three noun triggers “lack”, “absence”, and “freedom” as well as the adjective trigger “free” are considered for this type of negation. <b>Example:</b> Patients with good performance status,..., and the absence of systemic disease.	“lack”, “absence”, “freedom” “free”
<b>Verb Negation</b>	<b>Comment:</b> Some verbs can be used as negation triggers. We considered the verbs “deny”, “decline”, and “lack” as verb negation triggers. <b>Example:</b> Information on final patient outcomes was also lacking.	“deny” “decline” “lack”

In this step, the keyword of each negation class should be checked. For example, if there are some negation keywords in the reference annotation, the number of occurrences for each essential word should be the same in both the reference annotation and the learner annotation.

After the accomplishment of step 3, learner knowledge about the related concept can be judged and scored by “1” or “0.” If all 3 steps are passed successfully, the learner annotation is scored as “1”; otherwise it is scored “0.”

## Test and evaluation

To evaluate the proposed method, a virtual environment was designed to deliver an e-course and provide suitable tools for the learners to make their annotations. This virtual learning environment was implemented by means of required technologies such as ASP.Net 3.5, Microsoft access 2007 and C# programming language. The e-course was selected under the topic of “online education terms” in which the most pivotal online education terms were presented, defined, and discussed.

When the course is delivered to the learner and s/he starts the first stage of study by reading the course content. Afterward, the learner can visit the annotation page where the learning concepts are illustrated to the learner in a hierarchical tree. By clicking on each concept in the tree, the left side of related annotation appears and the learner can fill out the right side and save his/her annotation(s). After saving annotations, the learner is redirected to take a post-test consisting of multiple-choice questions on another page. On this page, question(s) about the same concept(s) are being presented to the learner and should be answered.

The implemented website has already saved 182 annotations of 45 learners. Learners were selected from BSc and MSc students in the field of Computer and Information Technology.

In another part of this work, three experts were asked to assess each annotation being made by the learners and determine the knowledge level of each learner regarding each concept. Experts were asked to give the score “1” to those annotations they detect correct and “0” to incorrect ones. After fulfilment of each part of the course, the learners were asked to take a multiple choice test and the obtained results were considered for comparison with annotation based assessment results. For every concept, at least one multiple choice question should be answered. The score for the correct answers is “1” and for any wrong answer is “0.” On the other hand, after completion of the annotation checking steps, the proposed method grades the knowledge of a learner about annotated concepts with “1” or “0.” (Figure 4) shows the learner assessment based on the evaluation of 20 annotations by the experts, the multiple choice test and the proposed method.

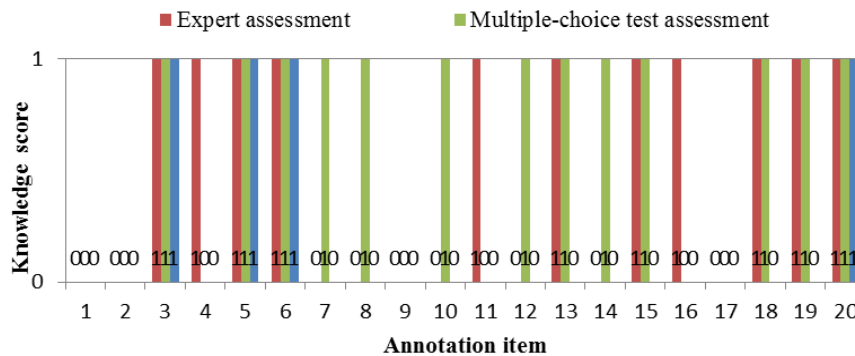


Figure4. Learner assessment for 20 annotations

As Figure 4 illustrates, in 8 items the results related to the knowledge evaluation being done by the experts, the multiple choice test, and the proposed method are the same, and in 13 items estimated knowledge by the experts and the proposed method are the same. This result considers 20 annotations, and now a True and False alarm rate metric is utilized to evaluate complete results for 182 items.

True and false alarm rate metric: The True alarm rate is utilized to measure the number of times that both the proposed method and human model scores are the same. On the other hand, the False alarm rate measures the number of times that the proposed method and human model scores are different (Ros'e et al., 2003). Figure 5 shows the true and false alarm rate of the experts and the multiple choice assessment. Figure 6 shows the true and false alarm rate of the proposed method and the experts' learner assessment.

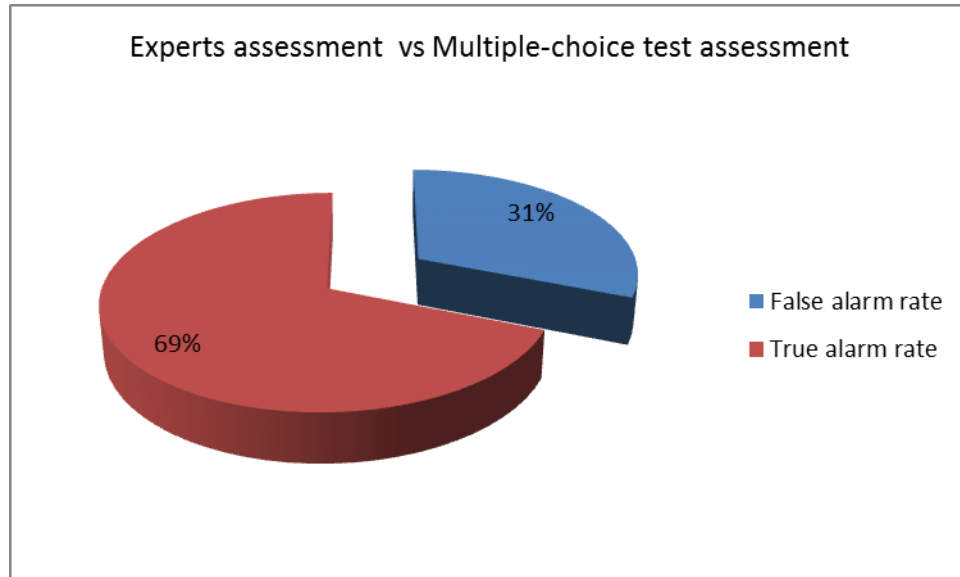


Figure5. Multiple choice test and experts assessment alarm rate

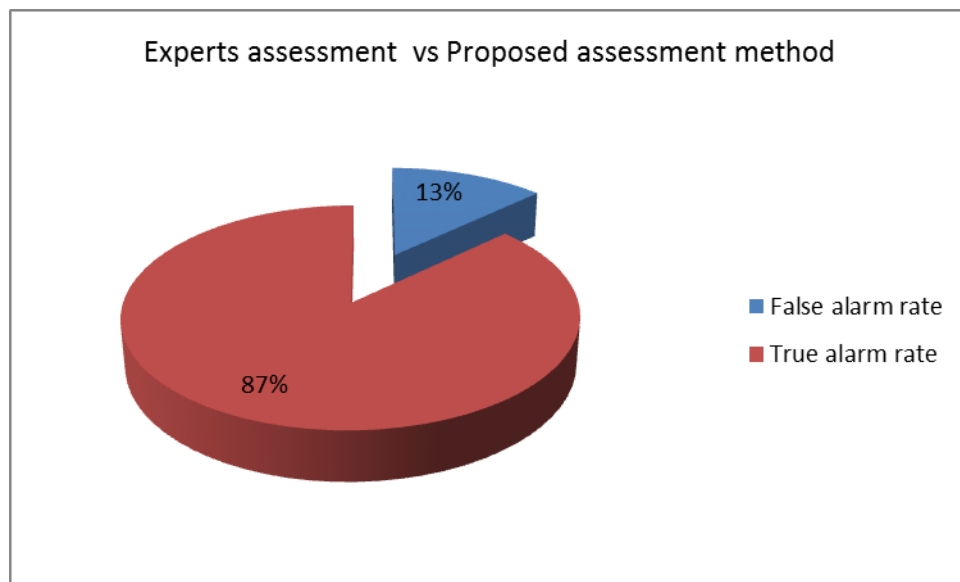


Figure6. Proposed method and experts' assessment alarm rate

As shown in Figure 5 and Figure 6, we reach a true alarm rate of 69% for assessed knowledge by experts vs. multiple choice test, and 87% for proposed method vs. experts. These are promising results for annotation-based assessment. It shows that False alarm rate for experts vs. multiple choice test assessment is greater than false alarm rate for proposed method vs. experts assessment. The true alarm rate between the multiple choice test and the experts' assessment is about 69% and the true alarm rate between the experts and the designed proposed method is 87% which shows an increase in accuracy of 18%. These results are reasonable because multiple choice tests cannot

assess learners accurately and therefore the results differ more from proposed method results, when compared to proposed method vs. experts' results differences.

The 87% true alarm rate for the proposed method vs. expert assessment shows that the 3 steps of annotation checking can successfully assess knowledge of learners and can be applied to automatic learner assessment.

In some cases, the proposed assessment method is different from expert assessment. For example, the learner annotation is very different from all reference annotations; however, the annotation is correct. This leads the annotation without essential words, to fail in the checking essential words step. To solve the problem, we should consider this user annotation as another reference annotation for this concept. If an annotation is considered correct by experts while the proposed method detection is incorrect, the annotation can be used as a new reference after refinement.

In some cases, expert detection is that the annotation is not correct but the proposed method considers it to be correct. This issue is due to the fact that the most similar reference annotation is found to be accidentally similar to the learner annotation in words order, contrast and essential words. That is why we used the  $\beta$  threshold. If  $Score(SRA)$  calculated by equation (9) is greater than  $\beta$ , then we consider this learner annotation for further processing. Using this threshold, the true alarm rate will become 89%.

## Conclusions

In this research, we proposed a new implicit assessment method based on learner's annotations. In this method, for each learning concept there are several reference annotations which are designed by human experts. The learner annotations are compared with related reference annotations which differ from each other in word choice, words order, and syntax. The M-BLEU algorithm was employed to choose the most similar reference annotation in comparison with the annotation being made by the learner. Afterward, final annotation checking is started in which the essential words, words order and contradiction of the learner annotation are checked and learner knowledge is estimated. We tested our method and showed that (comparing to the multiple choice test), the proposed method can increase the true alarm rate of assessment by about 20% (89% vs. 69%). It means that proposed method for learner's assessment is effective and can be useful for automatic assessment of learners.

The proposed annotation-based assessment has many applications. Although this method is applicable to free text answer assessment, we designed a framework to guide the user to annotate concepts and assess this annotation. This method has the benefit that the learner can be assessed implicitly. And the correct annotations can help users to remember concepts by reviewing their annotations. The best annotation also can be recommended to another learner when needed.

Another application is generating Automated Personalized Summaries based on assessed learner knowledge by eliminating contents that a learner already has its knowledge.

## References

- Belcadhi, L., Henze, N., & Braham, R. (2004). An assessment framework for elearning in the semantic web. In A. Abecker et al. (Eds.), *Proceedings of the Twelfth GI- Workshop on Adaptation and User Modelling in Interactive Systems* (pp. 11-16). Berlin, Germany: Humboldt-Universität.
- Brush, A. J. B. (2002). *Annotating digital documents for asynchronous collaboration* (Technical report no. 02-09-02). Seattle, WA: University of Washington.

- Burstein, J., Kukich, K., Wolff, S., Lu, C., Chodorow, M., Bradenharder, L., & Harris M. D. (1998). Automated scoring using a hybrid feature identification technique. *Proceedings of the Annual Meeting of the Association of Computational Linguistics* (pp. 206-210). doi: 10.3115/980451.980879
- Burstein, J., Leacock C., & Swartz, R. (2001). Automated evaluation of essay and short answers. In M. Danson (Ed.), *Proceedings of the Sixth International Computer Assisted Assessment Conference* (pp. 41-45). Loughborough, UK: Loughborough University.
- Callear, D., Jerrams-Smith, J., & Soh, V. (2001). Caa of short non-mcq answers. *Proceedings of the 5th International Computer Assisted Assessment conference*. Retrieved from <https://dspace.lboro.ac.uk/2134/1791>
- Christie, J. R. (1999). *Automated essay marking—for both style and content*. Retrieved from <https://www.cs.washington.edu/education/courses/cse590d/04sp/papers/christie-sear.pdf>
- Foltz, P., Laham, D., & Landauer, T. (1999). The intelligent essay assessor: Applications to educational technology. *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning*. Retrieved from <http://imej.wfu.edu/articles/1999/2/04/>
- Gindl, S., Kaiser, K., & Miksch, S. (2008). Syntactical negation detection in clinical practice guidelines. *Studies in Health Technology and Informatics*, 136, 187-192
- Larkey, L. S. (1998). Automatic essay grading using text categorization techniques. *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 90-95). doi: 10.1145/290941.290965
- Harabagiu, S., Hickl, A., & Lacatusu, V. F. (2006). Negation, contrast and contradiction in text processing. *American Association for Artificial Intelligence*, 1, 755-762.
- Mason, O., & Grove-Stephenson, I. (2002). Automated free text marking with paperless school. In M. Danson (Ed.), *Proceedings of the Sixth International Computer Assisted Assessment Conference* (pp. 213-95). Loughborough, UK: Loughborough University.
- Miller, G. A., Fellbaum, C., Teng, R., Wakefield, P., Poddar, R., & Langone, H. (2006). *WordNet: A lexical database for the English language*. Princeton, NJ: Princeton University Cognitive.
- Ming, Y., Mikhailov, A., & Kuan T. (2000). *Intelligent essay marking system, learners together*. Singapore: NgeeANN Polytechnic.
- Mitchell, T., Russell, T., Broomhead, P., & Aldridge, N. (2002). *Towards robust computerized marking of free-text responses*. In M. Danson (Ed.), *Proceedings of the Sixth International Computer Assisted Assessment Conference*. Loughborough, UK: Loughborough University.
- Noorbehbahani, F., & Kardan, A. A. (2010). The automatic assessment of free text answers using a modified BLEU algorithm. *Elsevier*, 56(2), 337-345.
- Otterbacher, J., & Radev, D. (2004). Comparing semantically related sentences: The case of paraphrase versus subsumption. *Proceedings of the 20th international conference on Computational Linguistics*. doi: 10.3115/1220355.1220539
- Page, E. B. (1996). Grading essay by computer: Why the controversy? *Hand-out for NCME Invited Symposium* (pp. 1-3). New York, NY: National Council on Measurement in Education.
- Papineni, K., Roukos, S., Ward, T., & Zhu, W. (2002). BLEU: A Method for automatic evaluation of machine translation. In P. Isabelle (Ed.), *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)* (pp. 311- 318). Philadelphia, Pennsylvania, USA: Association for Computational Linguistics.
- Petkovic, D., Lank, E., Ramirez, F. A., Raghavendra, S., Chen, F., Pekiner, C., ... Marquez, A. (2005). Asynchronous multimedia s for web-Based collaboration in biology education. In R. W. Lienhart, N. Babaguchi, & E. Y. Chang (Eds.), *Proceedings of SPIE Conference on Multimedia Storage and Retrieval* (pp.108-113). San Jose, California, United States: International Society for Optics and Photonics.
- Putri Ratna, A. (2006). Web based automated essay grading system using LSA method for Indonesian language with most important word. In E. Pearson & P. Bohman (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2006* (pp. 115-120). Chesapeake, VA: AACE.
- Rosé, C., Roque, A., & VanLehn, K. (2003). A hybrid text classification approach for analysis of student essays. *Processing of Build Educational Applications Using Natural Language* (pp. 68-75). doi: 10.3115/1118894.1118904
- Rudner, L., & Liang, T. (2002). Automated essay scoring using bayes' theorem. *The Journal of Technology, Learning, and Assessment*. Retrieved from <http://napoleon.bc.edu/ojs/index.php/jtla/article/view/1668>

Sukkarieh, J., Pulmand, S., & Raikes, N. (2003, October). *Auto-marking: using computational linguistics to score short, free text responses*. Paper presented at the 29<sup>th</sup> annual conference of the International Association for Educational Assessment (IAEA), Manchester, UK.

Vantage Learning Tech. (2000). *A study of expert scoring and intellimetric scoring accuracy for dimensional scoring of grade 11 student writing responses* (Report no. RB-397). Newtown, PA: Vantage Learning.

Webster, W. (1990). Computers and writing assessment: A preliminary view. *The international Journal of Computers and Composition*, 8(1), 7-22.

Whittington, D., & Hunt, H. (1999). Approaches to the computerized assessment of free text responses. In M. Danson (Ed.), *Proceedings of the Sixth International Computer Assisted Assessment Conference* (pp. 207-219). Loughborough, UK: Loughborough University.

## Fostering 5<sup>th</sup> Grade Students' Understanding of Science via Salience Analogical Reasoning in On-line and Classroom Learning Environments

Ming-Hua Chuang and Hsiao-Ching She\*

Institute of Education, National Chiao Tung University, 1001 Ta-Hsueh Rd, Hsin Chu City 300, Taiwan //  
scotte0809@msn.com // hcshe@mail.nctu.edu.tw

\* Corresponding author

(Submitted March 07, 2012; Revised August 07, 2012; Accepted January 21, 2013)

### ABSTRACT

This study examined whether or not on-line learning (on-line as compared to classroom) influenced the progression of scientific concept construction and analogical reasoning ability; and the degree to which different analogical reasoning (surface and relational salience) contributes to the progression of scientific concept construction and analogical reasoning ability. A total of 190 5<sup>th</sup> grade students were assigned to three different conditions: surface salience analogical reasoning without the use of on-line learning (condition 1), the same surface salience analogical reasoning with on-line learning (condition 2), and relational salience analogical reasoning with on-line learning (condition 3). Results indicate that condition 2 significantly outperformed condition 1, regardless of the progression of scientific concept construction and analogical reasoning. Condition 2 performed at the same level that condition 3 did, regardless of the progression of scientific concept construction and analogical reasoning. The stepwise regression indicated that the best predictor for scientific concept construction is analogical reasoning followed by scientific reasoning. This study highlights the idea that on-line learning benefits students more than classroom learning, surface salience analogical reasoning is as efficient as relational salience analogical reasoning, and analogical reasoning can better predict scientific concept construction than can scientific reasoning.

### Keywords

On-line learning vs. classroom learning, Relational salience vs. surface salience analogical reasoning, Image formation of eye

### Introduction

Analogy has been widely noted as being powerful for helping learners to make connections between the pre-existing base knowledge and new target (Driver & Bell, 1986) or to facilitate abstraction from their base knowledge to generate new schema (Duit, 1991) or to employ a combination of logical reasoning processes to assist in problem solving or inquiry (Nersessian, 1992). Many studies have shown that the use of analogy as an instructional approach would improve students' scientific concept learning (Podolefsky & Finkelstein, 2007; Zheng, Yang, Garica, & McCadden, 2008). Analogical reasoning also plays a prominent role in the field of scientific studies (Hofstadter, 2003). Analogical reasoning may even be “the main engine of inventive thinking” providing the ability to “generate new scientific knowledge, scientific model, and theory for interpretation of scientific phenomenon in scientific inquiry (Glynn, 1991; Hofstadter, 2003). The law of gravitation and atom model structure were developed by Newton and Rutherford-Bohr through the use of analogy to construct their new knowledge of scientific evidence (Dagher, 1998). These imply that analogical reasoning can be very powerful for instruction and learning, and particularly that it is a key skill for students to practice reasoning in learning science, develop their understanding of science, solve scientific problems, and accomplish scientific inquiry. These literatures highlight the prevalence of analogy in facilitating students' scientific conceptual learning and thinking, however there remains a lack of empirical studies to demonstrate how the use of analogy can facilitate students' analogical reasoning ability as time goes on.

### Literature

#### Analogical reasoning

Analogical reasoning essentially consists of encoding, inference, mapping, application and response (Sternberg, 1977). The Sternberg model focuses on conceptual inference in reasoning. Vosniadou (1989) addressed the idea that analogical reasoning was a process of mapping a known source (analog) system to an unknown target system in



order to facilitate the knowledge transfer, and further to make the correlation between the two systems. Gentner (1989) decomposed analogical reasoning into five sub-processes which are: access the base system, perform the mapping between base and target, evaluate the match, store the inferences in target, and extract the commonalities. Whereas Gentner's structure mapping approach predominately focuses on carry-over of prepositional structures, the pragmatic approach of Holyoak (1985) also takes into account contextual factors (Wilbers & Duit, 2001). Gentner views analogical reasoning as a comparison of similarities between base and target. Wilbers and Duit (2001) debates that both Gentner and Holyoaks consider propositionally based knowledge as a starting point for analogy use. As opposed to this, Wilbers and Duit (2001) proposed and claims that intuitive schema and mental models, spontaneously generated by the students when first confronted with the target phenomenon, are essential in analogy use. They further suggest that the analogy is a means of constructing (propositionally based) hypotheses on the basis of (image like) mental models and intuitive schemata triggered by the target phenomenon (Wilbers & Duit, 2001). They also claim that the process of analogy construction, which serves as heuristical exploration for the target, draws on a better known analog that provides some "proto-theory" for the yet unexplored target. The researchers agreed with their ideas that analogies are more of a tool to bring about hypotheses instead of proving them. The proof of hypotheses is a matter of empirical testing and beyond the use of analogy (Bunge, 1973).

Schonborn and Anderson (2008) proposed that it is very important to explicitly explain the relationships between analogy and target concepts to ensure that students interpreted the analogy as relational understanding instead of as a literal depiction of reality. They also addressed that both scientific knowledge and analogical reasoning ability are pivotal for determining whether or not an analogy will foster students' understanding (Schonborn & Anderson, 2008). Thus, analogical reasoning can help students to link the relationship between their own knowledge and new knowledge, and it can encourage students to think and repeatedly compare new and old concepts (May, Hammer, & Roy, 2006). Supported by constant thinking and reasoning, the analogical reasoning ability of students will refine and mature continuously.

The mechanism of analogy has been documented for more than a decade by many psychologists (Gentner, 1989; Vosniadou, 1989). Gentner (1989) studied analogy from similarity perspectives, and she considered analogy to have few attributes and many relations compared to metaphor which has few attributes and few relations. Her theory of analogical mapping emphasizes more on relational similarity than on surface similarity. She suggested that relational similarity is more critical for determining whether or not analogical mapping can be successful. Vosniadou (1989) defined analogical reasoning as a process of identifying, matching and transferring the structural information from base to target though the use of deductive, inductive and analogical reasoning processes. She proposed that objects with salient similarity can have either perceptual attributes or relational similarity and be accessed easily by learners. It remains unclear whether salient similarity with more perceptual attributes or with more relational similarity is more effective for students learning science concepts through the use of analogical reasoning. We are interested in exploring the relative effectiveness of relational salience and surface salience analogical reasoning on student learning. Thus, we specifically designed two different analogies, one being relational salience analogy and the other one being surface salience analogy. Relational salience analogy means that the analogy shares greater relational similarity and less surface (perceptual attributes) similarity with the target in comparison to the surface salience analogy. Surface salience analogy means that the analogy shares greater surface similarity (perceptual attributes) and less relational similarity with the target in comparison to the relational salience analogy.

### **Difficulty of scientific concept construction involving the image formation of eye**

The students' difficulties in learning about light and its role in image formation have been reported by many studies (Bendall, Goldberg, & Galili, 1993; Bouwens, 1987; Fetherstonhaugh, 1990). Their alternative perceptions were summarized in the following: (1) we see something just by light (visual rays), not by light being reflected to our eyes (Fetherstonhaugh, 1990; Shapiro, 1989); (2) the images are not caused by the reflection of light (Rice & Feher, 1987); (3) lenses are not necessary to form images (Fetherstonhaugh, 1990); (4) not sure what is happening between the object and our eyes to form an image with the light (Bendall, Goldberg, & Galili, 1993); (5) not sure about the relationships among eyes, object, and light sources during image formation (Selley, 1996); (6) confused about the refraction of convex lens and its image formation (Saxena, 1991).

A possible reason for students finding it so difficult to learn about light, vision, and image formation might be that these concepts are too abstract and they do not get an opportunity to manipulate and reproduce the process through

hands-on science. Though there are a few studies about helping students to develop their understanding of these concepts through different strategies (Fetherstonhaugh & Treagust, 1992; Selley, 1996), there is still little published evidence concerning analogical reasoning as a potential approach to reduce the difficulties of learning abstract concepts such as image formation of the eye. It would be of great interest to know whether or not analogical reasoning can facilitate students' learning of image formation of the eye.

### **On-line science learning**

In the past decade, on-line scientific learning activities have caught the attention of many science educators and researchers. Some studies have documented that on-line learning with well-developed science education theory benefits student learning more than conventional classroom learning does (She & Lee, 2008; Barak & Dori, 2005). However, Dantas and Kemm (2008) designed a physiology laboratory-based e-learning course to facilitate students' learning performance; they found there were no significant differences between the new course with e-learning and the conventional classroom learning. Cole and Todd (2003) suggest that merely transforming the materials from traditional paper-based to internet-based, without including advance designs in instruction content, may not produce better science learning performance.

On the other hand, others studies find that students assigned to a conventional course perform better than web-based course students in their fact-based science concepts (Hansen, Barnett, & MaKinster, 2004) and final exam (Wang & Newlin, 2000). Wang and Newlin (2000) further suggest that the reason why disagreement remains on whether on-line learning is more effective than conventional instruction on student attitudes and achievement, is that most on-line science learning programs do not include a well-developed educational theory as their design base. In order to resolve this problem, recent studies started to include well-developed educational theory into their on-line learning program. This resulted in better performance (Liao & She, 2009; Raes, Schellens, Wever, & Vanderhoven, 2012). However, there is a lack of studies regarding whether on-line learning outperforms the classroom instruction when they use the same well-developed science education theory such as salient analogical reasoning theory. Thus, we hope to explore whether on-line learning is more effective than classroom instruction when they all design their learning based upon the salience analogical reasoning theory. In addition, we are more interested in exploring whether surface and relational salience analogical reasoning on-line learning would bring the same success in terms of their scientific concept learning and analogical reasoning.

### **Purpose**

One of the aims of this study was to explore whether the use of on-line learning (on-line as compared to classroom) would influence the progression of scientific concept construction and analogical reasoning processes. A second aim was to investigate the way in which different analogical reasoning (surface and relational) contributes to the progression of scientific concept construction and analogical reasoning processes in mixed methods of analyzing test performance and interview results. In order to fulfil the aims, the following conditions were designed: in the first condition, students participated in surface salience analogical reasoning learning without an on-line aspect. The second and third conditions involve the on-line learning environment, but in slightly different ways. In the second condition, students participated in the same surface salience analogical reasoning learning but in an on-line learning environment. In the third condition (on-line learning environment), students took part in slightly different analogical reasoning - relational salience analogical reasoning - also with an on-line learning environment. The on-line learning groups in the second and third condition are referred to as on-line analogical reasoning groups. A third aim was to examine the relationships among scientific conceptions, scientific reasoning and analogical reasoning.

### **Methodology**

#### **Participants**

This study included 190 5th grade students (100 boys and 90 girls) from six classes of an urban public elementary school. The science achievement of these students was about average for 5<sup>th</sup> grade level at their school. Science was taught by the same science teacher across these six intact classes. Each class of students were always together for all

the courses in the same classroom, and they were mixed in heterogeneous level of achievement. Six intact classes of students were randomly assigned into three groups and each group with two intact classes of students. Before the experiment, the differences of pre-Scientific Conception Test (SCT), pre-Concept Dependent Analogical Reasoning Test (CDART), and pre-Scientific Reasoning Test (SRT) among three groups were examined to make sure there was no statistical significant difference among these three groups from condition 1 to 3. The mean and SD for each condition were provided in the Table 1. The ANOVA test showed that there was no statistically significant difference among three conditions, regardless of the performance of pre-SCT ( $F_{(2, 187)} = 1.57, p = .212$ ), pre-CDART ( $F_{(2, 187)} = 0.47, p = .624$ ), and pre-SRT ( $F_{(2, 187)} = 0.18, p = .839$ ).

## Procedure

The same content of image formation was covered for these three conditions. The difference of instruction for these three conditions were: Condition 1 arranged the presentation of surface salience analogy for learning the target concepts in classroom setting (34 boys and 31 girls), Condition 2 arranged the same presentation of surface salience analogy as condition 1 in the on-line learning environment (34 boys and 29 girls), Condition 3 arranged a slightly different analogical reasoning - relational salience analogy - for learning the target concepts in the on-line learning environment (33 boys and 29 girls). The instruction of each condition lasted an hour.

Pre-test of the Scientific Conception Test (SCT), Concept Dependent Analogical Reasoning Test (CDART), and the Scientific Reasoning Test (SRT) were administered to all students before the instruction in their regular classroom. A post-test and a retention-test of the SCT and the CDART were administered in their regular classroom after one- and seven-weeks instruction, respectively. In addition, 48 students from four classes (12 students from each class) of condition 2 and 3, who participated in the two slightly different on-line analogical reasoning groups, were interviewed and fully recorded by tape recorder before, one-week after, and 7 weeks after learning. Students were selected to interview based on the following basis: four high achievers (two boys and two girls), four middle achievers (two boys and two girls), and four low achievers (two boys and two girls). Five major questions (Appendix 1) were used to interview students' understanding and analogical reasoning involving the image formation of the eye for about fifteen minutes. Students' interview data were further analyzed according to the correct concepts and categories of analogical reasoning in order to determine whether the nature and extent of their scientific concepts and analogical reasoning ability improved as time went on.

## Learning materials

### *Design of on-line Learning Content: Image formation of Eye*

The design of the formation of eye learning materials was based upon the analogical reasoning theory of relational and surface salience analogy. A five-person team participated in the development of the image formation of eye learning materials: one science educator, two middle school science teachers, and two science education graduates.

The same content of image formation of eye and eye problems were covered by three groups which are (1) Structure of target (eye) and analogy (camera or transparent plastic eye model); (2) The mechanism for image formation of target (eye) and analogy (camera or transparent plastic eye model); (3) similarity and difference between image formation of the eye and camera/transparent plastic eye model (Figure 1 & Figure 2); (4) eye problem: the mechanism of causing nearsightedness and presbyopia and the way in which the eye can adjust for these problems with the use of analogy (Figure 3 & Figure 4). For condition 1, the science teacher used the transparent plastic eye model as the analogy to teach the image formation of the eye, and students interacted with the real model in the classroom setting. For condition 2, students learned the same analogy (transparent plastic eye model) through the on-line learning environment individually. For condition 3, students learned the image formation of eye using the camera as analogy in the on-line learning environment individually. In short, the difference between on-line and classroom instruction is that the on-line condition would visualize the image formation of the eye through on-line visualizing animation, and the classroom condition would learn the image formation of the eye through hands-on activities and lectures.

Research of Scientific concept e..x

esciedu.nctu.edu.tw/NeoSCORM/layout/new/popup.php?content=course&page=004069

Research of Scientific concept e..x



Course:Elementary Science and Life Technology

Section:Optics

Topic:Eyes (KSOL)

C-P00  
C-P01  
C-P02  
C-P03  
C-P04  
C-P05  
C-P06  
C-P07  
C-P08  
C-P09  
C-P10  
C-P11  
C-P12  
C-P13  
C-P14  
C-P15  
C-P16  
C-P17  
C-P18  
C-P19  
C-P20  
C-P21  
C-P22  
C-P23  
C-P24  
C-P25

### The similarity between camera and eye

Analogy	Target
<b>Camera</b> 	<b>Eye</b> 
<b>Camera lens</b>	<b>Eye lens</b>
<b>Film</b>	<b>Retina</b>
Camera lens gathers light	Eye lens gathers light
Image exposed on film =>shows clear image	Image falls on retina =>shows clear image
Smaller inverted image on the film	Smaller inverted image on the retina

Next page

Figure 1. The web-page of analogy and target concept comparison (for relational salience group)

Research of Scientific concept e..x

esciedu.scie.net/NeoSCORM/layout/new/popup.php?content=course&page=003996

Research of Scientific concept e..x

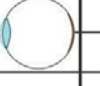

Course:Elementary Science and Life Technology

Section:Optics

Topic:Eyes (SSOL)

PA-p32  
PA-p0  
PA-p1  
PA-p2  
PA-p3  
PA-p4-1  
PA-p4-2  
PA-p5  
PA-p6-1  
PA-p6-2  
PA-p7  
PA-p8  
PA-p9  
PA-p10  
PA-p11-1  
PA-p11-2  
PA-p12

### The similarity between plastic eye ball and eye

Analogy	Target
<b>Plastic ball</b> 	<b>Eye</b> 
<b>Convex lens</b>	<b>Eye lens</b>
<b>Medium</b>	<b>Retina</b>
Convex lens gathers light	Eye lens gathers light
Image exposed on medium =>shows clear image	Image falls on retina =>shows clear image
Smaller inverted image on the medium	Smaller inverted image on the retina

Next page

Figure 2. The web-page of analogy and target concept comparison (for surface salience group)

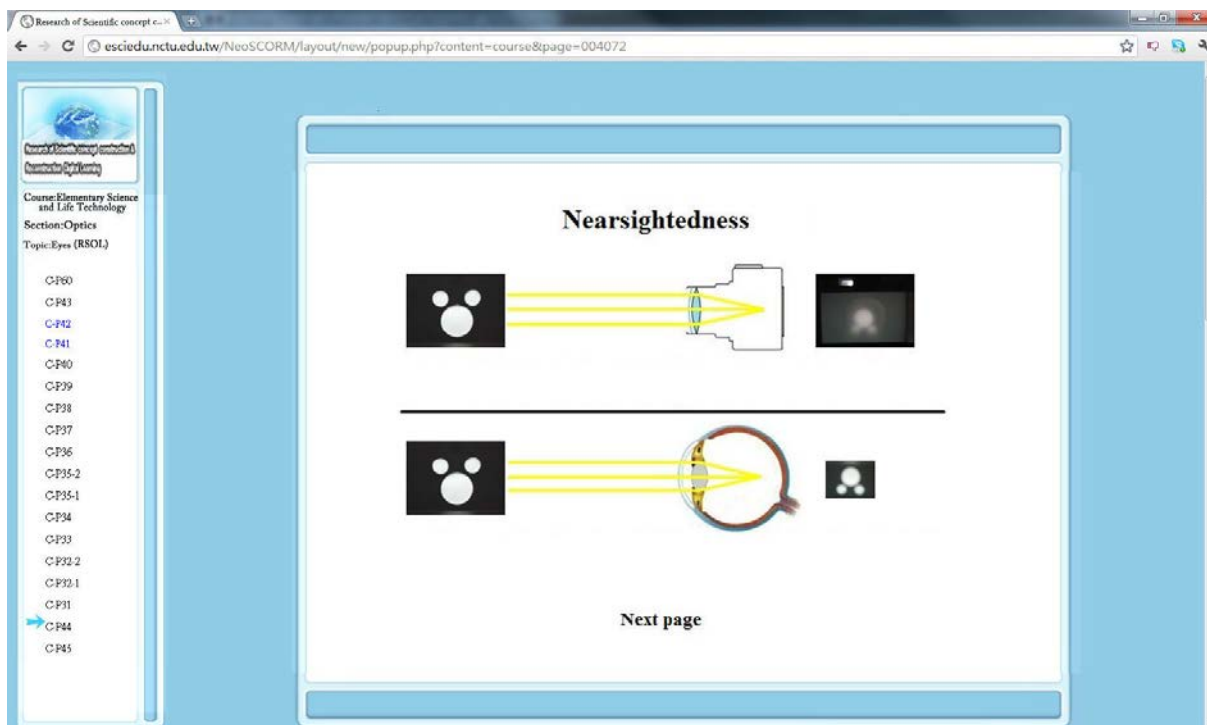


Figure 3. Web-page of the image formation mechanism of camera and eye (for relational salience group)

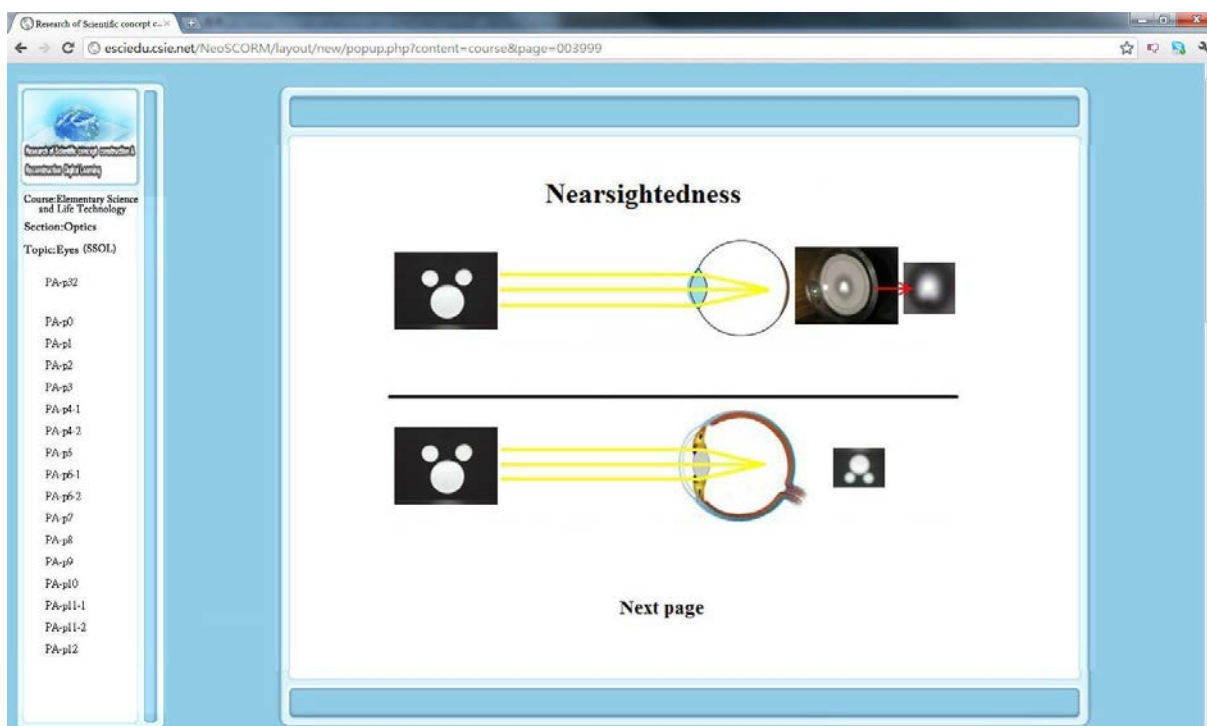


Figure 4. Web-page of the image formation mechanism of transparent plastic eye and eye (for surface salience group)

## Instruments

Scientific Conception Test (SCT), Concept Dependent Analogical Reasoning Test (CDART), and Scientific Reasoning Test (SRT)

The SCT and CDART were established by the same panel of five evaluators, ensuring that the items were properly constructed and relevant to the image formation of eye learning materials which students received. The modified SRT (Lawson, 1978) for elementary school students was used in the study (She & Lee, 2008). The SCT, CDART, and SRT all are the two-tier multiple-choice diagnostic instruments and each item contains two tiers. Students need to answer both tiers correctly in order to receive one point. There are 20, 13, and 12 items for SCT, CDART, and SRT, respectively. The pre-test, post-test, and retention-test of SCT and CDART are equivalent. The Cronbach  $\alpha$  for the pre-, post-, and retention-SCT were 0.69, 0.80, and 0.81, respectively. The Cronbach  $\alpha$  for the pre-, post-, and retention-CDART were 0.72, 0.80, and 0.78, respectively. The Cronbach  $\alpha$  of SRT was 0.71 which is close to the Lawson's result of Cronbach  $\alpha$  0.78.

The SCT was developed to measure students' image formation of eye related conceptions (Appendix 2), which required students to choose the answer and justify their response in the first and second tier of the question, respectively. These questions required students to use deeper information processing ability mainly involving analysis and synthesis. The CDART was developed to measure the degree of students' analogical reasoning involving image formation of the eye conceptions (Appendix 3), which required students to select the correct scientific conceptions and use their analogical reasoning in the first and second tier of the question, respectively. The SRT was developed to measure students' scientific reasoning abilities (included: deductive reasoning for aspects of conservation, proportional thinking, identification and control of variables, probabilistic thinking, correlative thinking, and hypothetical deductive), which required students to choose the correct answer and use the scientific reasoning abilities to make correct solution in the first and second tier of the question, respectively.

### Interview analysis

Students' interview results were transcribed and analyzed using the coding system developed by the researchers. We evaluated the nature and extent of students' scientific conception and analogical reasoning. Students' scientific conceptions were categorized according their correctness. About one fifth of the interviewee's results were checked by the second coder and inter-rater reliability was 0.90.

Three different categories were used to analyze the nature and extent of analogical reasoning, which were relational analogical reasoning, surface analogical reasoning, and relational reasoning. The inter-rater reliability was 0.93. The coding system used for analogical reasoning was as follows:

- Surface analogical reasoning: This focuses on the students' use of object attributes shared between the target and analogy to reason. For instance, "convex (camera) lens are like eye lens" or "the plastic ball (camera) is transparent (oblong shape), but the eye is not" were coded as surface analogical reasoning.
- Relational analogical reasoning: This focuses on the students' use of relation predicates shared between the target and analogy to reason. For instance, "convex (camera) lens are like eye lens; both of them converge the light and form an inverse real image" was coded as relational analogical reasoning.
- Relational reasoning: This focused on students' use of target object relations to reason without the use of analogy as support. The reasoning level is highest among the four of them, because the students have passed the process of analogy exploring, hypothesis testing and proving; they are beyond the use of analogy. For instance, "It would converge when light passes the lens of the eye, thus the image would form in the retina" was coded as relational reasoning.

## Results

### Multivariate analysis of the Scientific Conception Test (SCT)

The descriptive statistics and repeated measures of ANOVA were conducted to examine any increase from pre-, post- to retention-SCT for each condition (Table 1). This indicates that the increases in pre-, post- and retention-SCT mean scores also reached statistical significance, regardless of condition 1 ( $F_{(2, 128)} = 16.12^{***}$ ,  $p = .000$ ), condition 2 ( $F_{(2, 122)} = 54.87^{***}$ ,  $p = .000$ ) and condition 3 ( $F_{(2, 120)} = 55.97^{***}$ ,  $p = .000$ ). The post hoc test suggests that the post- and retention-SCT were significantly higher than pre-SCT across three conditions.

Table 1. Descriptive statistics of the pre-, post-, and retention of SCT, CDART, and SRT

	N	Pre-test		Post-test		Retention test			
		Mean	S.D.	Mean	S.D.	Mean	S.D.	<i>F</i>	<i>p</i>
<i>Scientific Concept Test (SCT)</i>									
Condition 1	65	7.38	3.05	8.69	3.05	9.83	4.38	16.12***	.000
Condition 2	63	6.94	3.12	10.48	4.72	11.35	4.48	54.87***	.000
Condition 3	62	6.37	3.52	10.56	4.65	11.39	3.89	55.97***	.000
<i>Concept Dependent Analogical Reasoning Test (CDART)</i>									
Condition 1	65	4.77	2.37	5.62	2.93	6.86	2.46	22.03***	.000
Condition 2	63	4.95	2.78	7.40	3.00	7.53	3.18	28.66***	.000
Condition 3	62	4.48	2.97	6.23	3.01	6.98	2.80	20.23***	.000
<i>Scientific Reasoning Test (SRT)</i>									
Condition 1	65	2.88	1.56						
Condition 2	63	2.92	2.03						
Condition 3	62	2.74	1.64						

Condition 1: surface salience classroom group, Condition 2: surface salience on-line group, Condition 3: Relational salience on-line group.

The first part was to investigate whether the on-line learning environment or the classroom setting would influence the progression of scientific concept more with the use of surface salience analogy. The Levene test of homogeneity was not significant on pre-SCT ( $F_{(1, 125)} = 0.11, p = .743$ ), thus one-factor MANCOVA was conducted to examine the effects of instructional approaches using post- and retention-SCT scores as the dependent measures, and students' pre-SCT scores as the covariate. The results indicated that instructional approaches (Wilk's  $\Lambda = 0.89, F_{(2, 123)} = 7.65^{**}, p = .001$ ) reached a statistically significant effect on the performance of post- and retention-SCT (Table 2). Therefore, the following main effect for instructional approach was performed. The main effect was performed to independently examine the effect of the instructional approaches on post- and retention-SCT. This indicated that the effects for instructional approaches were significant on both post-SCT scores ( $F_{(1, 124)} = 13.48^{***}, p = .000$ ) and retention-SCT ( $F_{(1, 124)} = 8.91^{**}, p = .003$ ). The post-hoc analysis for main effect suggests that the condition 2 groups performed significantly better than the condition 1 group (on-line > classroom,  $p_{(post)} = .000, p_{(retention)} = .003$ ) on their post- and retention-SCT.

Table 2. MANCOVA results of SCT for on-line and classroom groups

Source of variance	Wilk's $\Lambda$	Multivariate $F$	Univariate $F$		Post-hoc
			Post-test	Retention-test	
<i>Covariates</i>					
Pre-test scores	0.59	43.51*** (.000)			
<i>Group membership</i>					
Instructional approaches	0.89	7.65** (.001)	13.48*** (.000)	8.91** (.003)	Post: Condition 2 > Condition 1 (.000) Retention: Condition 2 > Condition 1 (.003)

\*\*\* $p < .001$ , \*\* $p < .01$ , \* $p < .05$ .

SCT: scientific conception test.

Condition 1: surface salience classroom group, Condition 2: surface salience on-line group.

Wilks' lambda is most commonly used statistic for overall significance. It considers differences over all the characteristic roots.

Wilks' lambda is used to examine the relationship between instructional approaches and overall performance on SCT. Wilks' lambda ranges from 0 to 1, and the lower the Wilks' lambda the more the given effect contributes to the model.

The second part was to examine the way in which different analogical reasoning (surface and relational salience) on-line learning contributes to the progression of scientific concepts. The Levene test of homogeneity was not significant on pre-SCT ( $F_{(1, 121)} = 2.97, p = .087$ ), thus one-factor MANCOVA was conducted. However, there was no statistically significant difference between condition 2 and condition 3 in terms of their post- and retention-SCT (Wilk's  $\Lambda = 0.99, F_{(2, 119)} = 0.36, p = .700$ ).

## Multivariate analysis of Concept Dependent Analogical Reasoning Test (CDART)

The descriptive statistics and repeated measures of ANOVA were conducted to examine any increase from pre-, post- to retention-CDART (Table 1) for each condition. The result showed an increase in pre-, post- to retention-CDART that also reached statistical significance, regardless of condition 1 ( $F_{(2, 126)} = 22.03^{***}$ ,  $p = .000$ ), condition 2 ( $F_{(2, 122)} = 28.66^{***}$ ,  $p = .000$ ) and condition 3 ( $F_{(2, 120)} = 20.23^{***}$ ,  $p = .000$ ). The post hoc test suggests that the post- and retention-CDART were significantly higher than pre-CDART for all conditions.

The first part was to investigate whether the on-line learning environment or the classroom setting would have a greater influence on the progression of scientific analogical reasoning when they both use the same surface salience analogy. The Levene test of homogeneity was not significant on pre-CDART ( $F_{(1, 124)} = 1.52$ ,  $p = .220$ ), thus one-factor MANCOVA was conducted to examine the effects of instructional approaches using post- and retention-CDART scores as the dependent measures, and students' pre-CDART scores as the covariate. The results indicated that instructional approaches (Wilk's  $\Lambda = 0.90$ ,  $F_{(2, 122)} = 6.60^{**}$ ,  $p = .002$ ) had a statistically significant effect on their performance of post- and retention-CDART (Table 3). Therefore, the following main effect for instructional approach was performed. The univariate F (one-factor ANCOVA) was performed to examine the effect of instructional approaches on post- and retention-CDART independently. It indicated that the effects for instructional approach were significant on post-CDART scores ( $F_{(1, 123)} = 12.57^{**}$ ,  $p = .001$ ), and not significant on retention-CDART. The post-hoc indicated that the students in condition 2 performed significantly better than did the condition 1 ( $p_{(post)} = .001$ ) on their post-CDART.

Table 3. MANCOVA results of CDART for on-line and classroom groups

Source of variance	Wilk's $\Lambda$	Multivariate $F$	Univariate $F$		Post-hoc
			Post-test	Retention-test	
<i>Covariates</i>					
Pre-test scores	0.75	20.29*** (.000)			
<i>Group membership</i>					
Instructional approaches	0.90	6.60** (.002)	12.57** (.001)	1.51 (.222)	Post: Condition 2 > Condition 1 (.001)

\*\*\* $p < .001$ , \*\* $p < .01$ , \* $p < .05$ .

CDART: concept dependent analogical reasoning test.

Condition 1: surface salience classroom group, Condition 2: surface salience on-line group.

The second part was to examine the way in which different analogical reasoning (surface and relational salience) on-line learning contributes to the progression of analogical reasoning. The Levene test of homogeneity was not significant on pre-CDART ( $F_{(1, 121)} = 0.89$ ,  $p = .346$ ), thus one-factor MANCOVA was conducted. However, there was no statistically significant difference between condition 2 and condition 3 in terms of their post- and retention-CDART (Wilk's  $\Lambda = 0.97$ ,  $F_{(2, 119)} = 2.15$ ,  $p = .121$ ).

## Stepwise regression analysis

This section examines whether students' degree of conceptual construction would be impacted by their analogical reasoning ability or by scientific reasoning ability. Therefore, the stepwise regression method was used to explore whether the pre-CDART or the pre-SRT would be most important for predicting the post-SCT scores, and whether pre-CDART or the pre-SRT would be the most important factor for predicting the retention-SCT scores.

Results indicated that the best predictor for post-SCT scores was the pre-CDART followed by pre-SRT scores. The standardized regression coefficient for pre-CDART and pre-SRT were 0.33 and 0.30. Together, pre-CDART and pre-SRT accounted for 24.1% of the variance in post-SCT scores and reached a statistical significance level ( $R^2 = .24$ ,  $F_{(2, 186)} = 29.53^{***}$ ,  $p = .000$ ). Results also showed that the best predictor for retention-SCT scores was pre-CDART followed by pre-SRT scores. The standardized regression coefficient for pre-CDART and pre-SRT were 0.38 and 0.30. Together, pre-CDART and pre-SRT accounted for 28.7% of the variance in retention-SCT scores and reached a statistical significance level ( $R^2 = .29$ ,  $F_{(2, 186)} = 37.55^{***}$ ,  $p = .000$ ) (Table 4).



Table 4. Stepwise multiple regression among SCT, CDART, and SRT

Significant Predictor Variable	Standardized Regression Coefficients	<i>t</i>	<i>p</i>	Cumulative % of Variance Explained
Post-test of Scientific Concept Test				
Pre-test of Concept Dependent Analogical Reasoning Test	0.33	4.95***	.000	15.8%
Pre-test of Scientific Reasoning Test	0.30	4.50***	.000	24.1%
Multiple R	0.49			
R <sup>2</sup>	0.24			
Retention-test of Scientific Concept Test				
Pre-test of Concept Dependent Analogical Reasoning Test	0.38	5.94***	.000	20.4%
Pre-test of Scientific Reasoning Test	0.30	4.66***	.000	28.7%
Multiple R	0.54			
R <sup>2</sup>	0.29			

\*\*\* $p < .001$ , \*\* $p < .01$ , \* $p < .05$ .

### Analysis of interview

The students in condition 2 and 3 were given five interview questions regarding the conceptions of image formation before, one week after and seven weeks after instruction, and the results were transcribed and further analyzed according to the quantity of correct concepts and level of analogical reasoning.

### Quantity of correct concepts

The quantity of correct concepts of interview results from before, one week after, and seven weeks after learning were analyzed. One-factor repeated measures of ANOVA were conducted to examine the effects of instructional approaches and any increase from pre-, post- to retention-correct concepts, respectively. The result indicated that instructional approaches did not reach a statistically significant effect on the performance of pre-, post- to retention-correct concepts. However, it indicated that the increases in pre-, post- to retention-mean score of correct concepts reached statistical significance ( $F_{(2, 90)} = 53.62^{***}$ ,  $p = .000$ ). The Mauchly's test of sphericity was significant at 0.05, so the Huynh-Feldt of  $F$  was used. The post hoc test suggests that the mean scores of correct concepts in the post- and retention-interview results were significantly higher than pre-interview results ( $p = .000$ ,  $p = .000$ ) (Table 5). Results indicated that students' mean scores of correct concepts significantly progressed from pre- to post-interview and pre- to retention-interview after they received the learning program regardless of condition 3 and condition 2 (Figure 5).

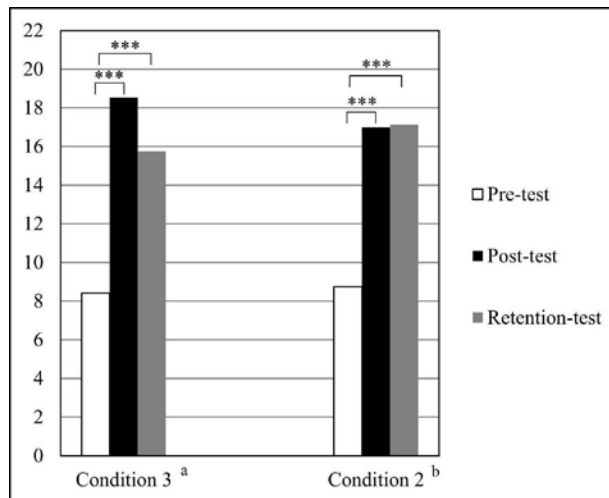


Figure 5. Distribution of students' correct concepts for interview questions

\*\*\* $p < .001$ , \*\* $p < .01$ , \* $p < .05$ .

<sup>a</sup> Condition 3: relational salience on-line group, <sup>b</sup> Condition 2: surface salience on-line group.

Table 5. One-factor repeated measures ANOVA results of students' correct concepts (for interview questions)

Source of variance	N	F	p	Post hoc
Instructional approaches	47	0.05	.818	
Time	47	53.62***	.000	post>pre(.000), retention>pre(.000)
Instructional approaches × Time	47	0.79	.456	

\*\*\* $p < .001$ , \*\* $p < .01$ , \* $p < .05$ .

### Quantity of analogical reasoning

Three categories of analogical reasoning: surface analogical reasoning (SAR), relation analogical reasoning (RAR), and relational reasoning (RR) were used to measure the quantity of analogical reasoning in which students made progress from pre-, to post-, and then to retention. The descriptive statistic of surface analogical reasoning, relational analogical reasoning, and relational reasoning for students who are in condition 2 and 3 are in Figure 6.

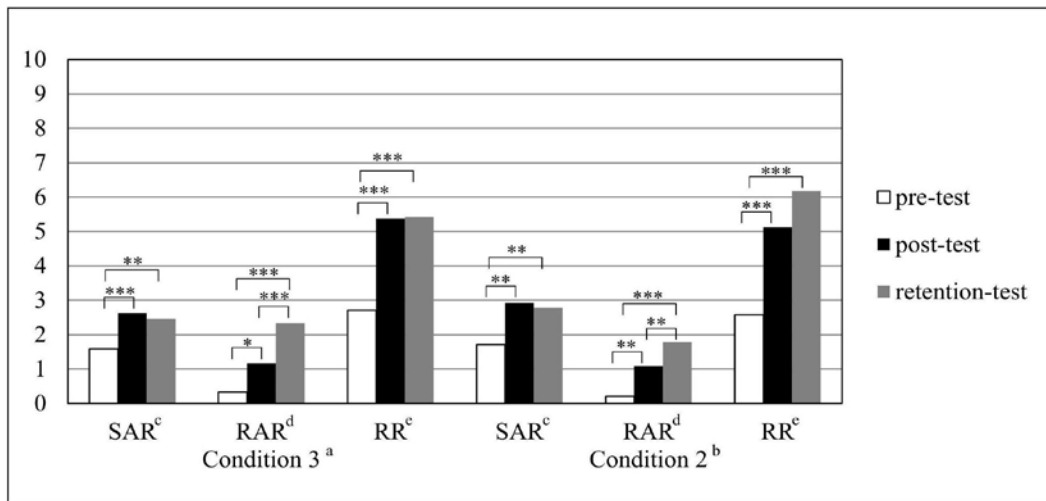


Figure 6. Distribution of students' analogical reasoning for interview questions

\*\*\* $p < .001$ . \*\* $p < .01$ . \* $p < .05$ .

<sup>a</sup> Condition 3: relational salience on-line group, <sup>b</sup> Condition 2: surface salience on-line group.

<sup>c</sup> SAR: surface analogical reasoning, <sup>d</sup> RAR: relational analogical reasoning, <sup>e</sup> RR: relational reasoning.

One-factor repeated measures of ANOVA were conducted to examine the effects of instructional approaches and any increase from pre-, post- to retention-SAR, RAR, and RR, respectively. The results indicated that instructional approaches did not reach a statistically significant effect on the performance of pre-, post- to retention-SAR, RAR, and RR, respectively. However, results indicate that the increases in pre-, post- to retention-mean score of SAR, RAR, and RR all reached a statistical significance level ( $F_{(2, 90)}(\text{SAR}) = 29.20^{***}$ ,  $p = .000$ ;  $F_{(2, 90)}(\text{RAR}) = 54.96^{***}$ ,  $p = .000$ ;  $F_{(2, 90)}(\text{RR}) = 60.35^{***}$ ,  $p = .000$ ). The Mauchly's test of sphericity was not significant at 0.05, so the sphericity of F was used. The post hoc test suggests that the mean scores of SAR in post- and retention- interview results were significantly higher than pre-interview results ( $p = .000$ ,  $p = .000$ ), regardless of condition 2 or condition 3. A similar pattern was found for the dimensions of RR ( $p = .000$ ,  $p = .000$ ), regardless of condition 3 or condition 2. For the dimensions of RAR, the mean scores of RAR in post- and retention-interview results were significantly higher than pre-interview results ( $p = .000$ ,  $p = .000$ ), and retention-interview results were significantly higher than post-interview results ( $p = .000$ ), regardless of condition 3 or condition 2 (Table 6).

Table 6. One-factor repeated measures ANOVA results of students' analogical reasoning (for interview questions)

Source of variance	N	F	p	Post hoc
Surface Analogical Reasoning				
Instructional approaches	47	1.43	.238	
Time	47	29.20***	.000	post > pre (.000), retention > pre .000)
Instructional approaches × Time	47	0.20	.824	

Relational Analogical Reasoning				
Instructional approaches	47	1.29	.262	
Time	47	54.96***	.000	post > pre (.000), retention > pre (.000), retention > post (.000)
Instructional approaches × Time	47	1.33	.271	
Relational Reasoning				
Instructional approaches	47	0.20	.658	
Time	47	60.35***	.000	post > pre (.000), retention > pre (.000)
Instructional approaches × Time	47	1.08	.343	

\*\*\* $p < .001$ , \*\* $p < .01$ , \* $p < .05$ .

## Discussion and conclusions

This study documented the findings that the surface salience analogical reasoning on-line learning group (condition 2) significantly outperformed the surface salience analogical reasoning classroom learning group (condition 1) on their students' post-test and retention-test of scientific concept construction and post-test of content dependent analogical reasoning (CDRT). The retention-CDRT did not reach a statistically significant difference level which may indicate that condition 2 has greater immediate instead of retained impact on students' analogical reasoning than does condition 1. Previous studies have demonstrated that classroom instruction with a well-developed science learning theory of surface analogical reasoning is better than conventional instruction without including the surface analogical reasoning theory (Chiu & Lin, 2005; Sarantopoulos & Tsapalis, 2004), which highlighted the idea that surface analogical reasoning is efficient for promoting student learning in a regular classroom setting. This study moves one step beyond with the demonstration that the on-line learning group outperformed the classroom learning group when they both used the same theories of surface salience analogical reasoning on their scientific concept construction and analogical reasoning ability.

The second focus of this study is to examine any differences existing between relational and surface salience in two different on-line analogical reasoning programs in terms of their analogical reasoning ability and scientific concept construction performance. Our mixed method demonstrated that there were no significant differences in the scientific concept construction and analogical reasoning between the relational salience analogical on-line learning group (condition 3) and the surface salience analogical on-line learning group (condition 2). The result did not support the idea that relational similarity is more critical for deciding if analogical mapping can be successful, which was proposed by Gentner (1989). Vosniadou (1989) proposed that objects with salient similarity can have either perceptual attributes or relational similarity and be accessed easily by learners. Our results seem more likely to support Vosniadou's (1989) ideas that the object with salient similarity can have either perceptual attributes or relational similarity and be easily accessed by learners. Our results move one step beyond to claim that both relational salience and surface salience analogy can have the same efficiency for promoting students' scientific concepts construction and analogical reasoning.

In addition, this study also provides further important information regarding the progression of students' scientific concept construction and analogical reasoning. Qualitative and quantitative results showed that both groups made significant progress from pre- to post- and pre- to retention-performance of concept construction and analogical reasoning. It demonstrated that students' scientific concept construction and analogical reasoning can be promoted significantly over time, regardless of employing surface or relational salience analogical reasoning. It sheds light on the idea that students can benefit through the use of analogical reasoning if the properties of analogy are salient to them. It does not necessarily have to be restricted to relation or surface to be successful.

Our stepwise regression showed other interesting results in that the best single predictor for scientific concept construction is analogical reasoning followed by scientific reasoning. Wilbers and Duit (2001) suggested that analogy is a means of constructing (propositionally based) hypotheses on the basis of (image like) mental models and intuitive schemata triggered by the target phenomenon. The researchers agreed with their ideas that analogy is a tool

to bring about hypotheses and find proof for that hypothesis. Other researchers also proposed that analogical reasoning can help students to link the relationship between their own knowledge and new knowledge as well as encourage students to think and repeatedly compare the new concepts and the old concepts (May, Hammer, & Roy, 2006). These explanations provided above support our result that analogical reasoning is a better predictor than scientific reasoning for scientific concept construction in the learning of image formation of the eye.

Analogy has been reported as an effective method to reduce the degree of difficulty of acquiring invisible or abstract scientific concepts (She, 2004; Dagher, 1998). Vosniadou (1989) suggested that analogical reasoning is used to map a known source (analog) system to an unknown target system in order to make a relationship between the systems and to further facilitate knowledge transfer. It provides insight that including analogical reasoning into science teaching can efficiently foster students' concept construction. The camera was the most widely used analogy in the majority science textbook to teach optical concepts about the formation of image and structure of the eye (Glynn, Britton, Semrud-Clikeman, & Muth, 1989; Iding, 1997). This study opens a new window that using a transparent plastic eye model is about as effective for helping students acquire the concepts of image formation and structure of the eye as using the camera. Gentner (1989) proposed that learners need to efficiently encode information at working memory and retrieve information from long-term memory to generate a potential candidate analogy. Once the degree of matching candidate analogy with target reached the appropriate level, analogical reasoning is complete (Gentner, 1989). Researchers have suggested that the use of students' familiar analogy would promote students' analogical reasoning ability (May, Hammer, & Roy, 2006). Our study indicated that the transparent plastic eye model is rather familiar with learners.

Moreover, this study indeed supports findings that the use of on-line learning with analogical reasoning theory fosters better scientific concept understanding and analogical reasoning ability than the classroom setting with the use of the same learning theory. One of the advantages of on-line learning is the easy inclusion of multimedia presentation, which provides students with opportunities to learn science through authentic learning (Miller & Miller, 2000). The design of our surface and relational salience analogical on-line learning was to provide students with authentic science learning opportunity through on-line multimedia presentation. Both surface and relational salience analogical reasoning on-line learning enhance students' scientific concept understanding and analogical reasoning ability with the same effectiveness.

## References

- Barak, M., & Dori, Y. J. (2005). Enhancing undergraduate students' chemistry understanding through project-based learning in an IT environment. *Science Education*, 89(1), 117–139.
- Bendall, S., Goldberg, F., & Galili, I. (1993). Prospective elementary teachers' prior knowledge about light. *Journal of Research in Science Teaching*, 30(9), 1169-1187.
- Bouwens, F. E. A. (1987). Misconceptions among pupils regarding geometrical optics. In J. D. Novak (Ed.), *Proceedings of Second International Seminar: Misconceptions and Educational Strategies in Science and Mathematics* (Vol. III, pp. 23-28). Ithaca, NY: Cornell University
- Bunge, M. A. (1973). *Method, model, and matter*. Boston, MA: Reidel Publishing Co.
- Chiu, M. H., & Lin, J. W. (2005). Promoting fourth graders' conceptual change of their understanding of electric current via multiple analogies. *Journal of Research in Science Teaching*, 42(4), 429-464.
- Cole, R. S., & Todd, J. B. (2003). Effects of web-based multimedia homework with immediate rich feedback on student learning in general chemistry. *Journal of Chemical Education*, 80(11), 1338–1343.
- Dagher, Z. R. (1998). The case for analogies in teaching science for understanding. In J. J. Mintzes, J. H. Wandersee, & J. D. Novak (Eds.), *Teaching science for understanding: A human constructivist view* (pp. 195-211). San Diego, CA: Academic Press.
- Dantas, A. M., & Kemm, R. E. (2008). A blended approach to active learning in a physiology laboratory-based subject facilitated by an e-learning component. *Advances in Physiology Education*, 32(1), 65–75.
- Driver, R., & Bell, B. (1986). Students' thinking and the learning of science: a constructivist view. *School Science Review*, 67, 443-356.
- Duit, R. (1991). On the role of analogies and metaphors in learning science. *Science Education*, 75(6), 649-672.

- Fetherstonhaugh, T. (1990). Misconceptions and light: a curriculum approach. *Research in Science Education*, 20(1), 105-113.
- Fetherstonhaugh, T., & Treagust, D. F. (1992). Students' understanding of light and its properties: Teaching to engender conceptual change. *Science Education*, 76(6), 653-672.
- Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp.199-241). New York, NY: Cambridge University Press.
- Glynn, S. M. (1991). Explaining science concepts: A teaching-with-analogies model. In S. M. Glynn, R. H. Yeany, & B. K. Britton (Eds.), *The psychology of learning science* (pp. 219-240). Hillsdale, NJ: Lawrence Erlbaum.
- Glynn, S. M., Britton, B. K., Semrud-Clikeman, M., & Muth, K. D. (1989). In J. A. Glover, R. R. Ronning, & C. R. Reynolds (Eds.), *Analogical reasoning and problem solving in science textbooks, Handbook of creativity* (pp.383-398). New York, NY: Plenum Press.
- Hansen, J. A., Barnett, M., & MaKinster, J. G. (2004). The impact of three-dimensional computational modeling on student understanding of astronomical concepts: A quantitative analysis. *International Journal of Science Education*, 26(11), 1365-1378.
- Hofstadter, D. (2003, May). Analogy as the central motor of discovery in physics. Paper presented at a Physics Department Colloquium, Ohio State University, Columbus, Ohio.
- Holyoak, K. J. (1985). The pragmatics of analogical transfer. *The Psychology of Learning and Motivation*, 19, 59-87.
- Iding, M. K. (1997). How analogies foster learning from science texts. *Instructional Science*, 25(4), 233-253.
- Lawson, A. E. (1978). Development and validation of the classroom test of formal reasoning. *Journal of Research in Science Teaching*, 15(1), 11-24.
- Liao, Y. W., & She, H. C. (2009). Enhancing eight grade students' scientific conceptual change and scientific reasoning through a web-based learning program. *Educational Technology & Society*, 12(4), 228-240.
- May, D. B., Hammer, D., & Roy, P. (2006). Children's analogical reasoning in a third-grade science discussion. *Science Education*, 90(2), 316-330.
- Miller, S. M., & Miller, K. L. (2000). Theoretical and practical constructions in the design of web-based instruction. In B. Abbey (Ed.), *Instructional and cognitive impacts of web-based education* (pp. 156-177). Hershey: Idea Group Publishing.
- Nersessian, N. J. (1992). How do scientists think? Capturing the dynamics of conceptual change in science. In R. N. Giere (Ed.), *Cognitive models of science* (Vol. XV, pp. 3-44.) Minnesota studies in the philosophy of science. Minneapolis, MN: University of Minnesota press.
- Podolefsky, N. S., & Finkelstein, N. D. (2007). Analogical scaffolding and the learning of abstract ideas in physics: An example from electromagnetic waves. *Physical Review Special Topics- Physics Education Research*, 3(1), 1-12.
- Raes, A., Schellens, T., Wever, B. D., & Vanderhoven, E. (2012). Scaffolding information problem solving in web-based collaborative inquiry learning. *Computers & Education*, 59(1), 82-94.
- Rice, K., & Feher, E. (1987). Pinholes and images: Children's conceptions of light and vision I. *Science Education*, 71(4), 629-639.
- Sarantopoulos, P., & Tsapalis, G. (2004). Analogies in chemistry teaching as a means of attainment of cognitive and affective objectives: A longitudinal study in a naturalistic setting, using analogies with a strong social content. *Methods and Issues of Teaching and Learning*, 5(1), 33-50.
- Schonborn, K. J., & Anderson, T. R. (2008). Bridging the educational research-teaching practice gap- Conceptual understanding part 2- Assessing and developing student knowledge. *Biochemistry and Molecular Biology Education*, 36(5), 372-379.
- Selley, N. J. (1996). Towards a phenomenography of light and vision. *International Journal of Science Education*, 18(7), 837-846.
- Shapiro, B. L. (1989). What children bring to light: Giving high status to learners' views and action in science. *Science Education*, 73(6), 711-733.
- She, H. C. (2004). Facilitating changes in ninth grade students' understanding of dissolution and diffusion through DSLM instruction. *Research in Science Education*, 34(4), 503-525.
- She, H. C., & Lee, C. Q. (2008). SCCR digital learning system for scientific conceptual change and scientific reasoning. *Computers & Education*, 51(2), 724-742.
- Sternberg, R. J. (1977). Component processes in analogical reasoning. *Psychological Review*, 84(4), 353-378.

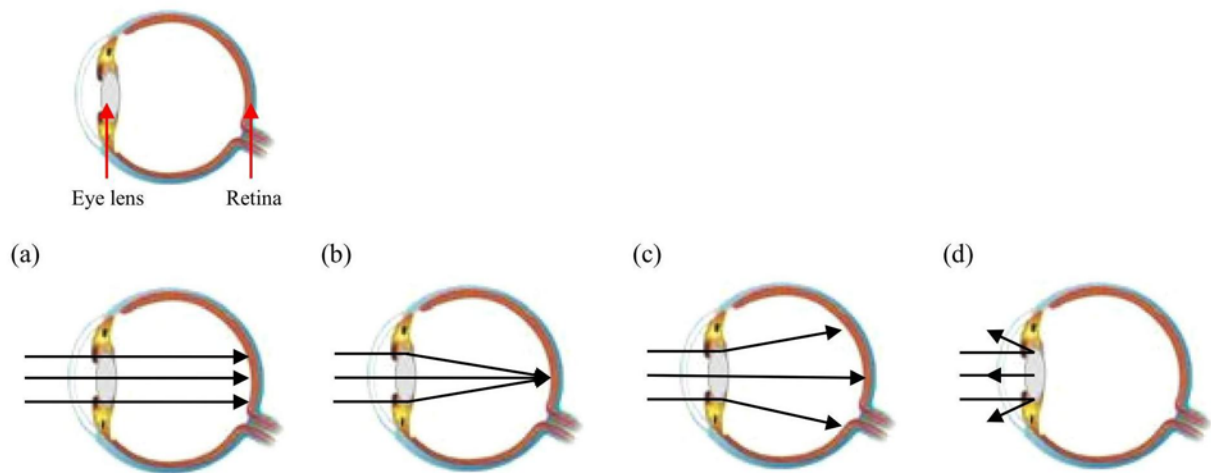
- Vosniadou, S. (1989). Analogical reasoning as a mechanism in knowledge acquisition: A developmental perspective. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 413-437). Cambridge: Cambridge University Press.
- Wang, A. Y., & Newlin, M. H. (2000). Characteristics of students who enroll and succeed in psychology Web-based classes. *Journal of Educational Psychology*, 92(1), 137-143.
- Wilbers, J., & Duit, R. (2001). On the micro-structure of analogical reasoning: The case of understanding chaotic systems. *Research in Science Education- Past, Present, and Future*, 10, 205-210.
- Zheng, R. Z., Yang, W., Garcia, D., & McCadden, E. P. (2008). Effects of multimedia and schema induced analogical reasoning on science learning. *Journal of Computer Assisted Learning*, 24, 474-482.

## Appendix 1. Interview Questions

- (1) Please describe the similarities and differences between the analogy of the plastic ball/camera and the target of eye.
- (2) How can we see objects? How does the eye work?
- (3) What happens when light passes through the convex lens and the concave lens?
- (4) What is nearsightedness? Why is that people with nearsightedness cannot see objects clearly?
- (5) What is\ presbyopia? Why is that people with presbyopia cannot see objects clearly?

## Appendix 2. Scientific Conception Test (SCT) Example Question

Please choose the correct answer regarding the pathway of light through the eye.

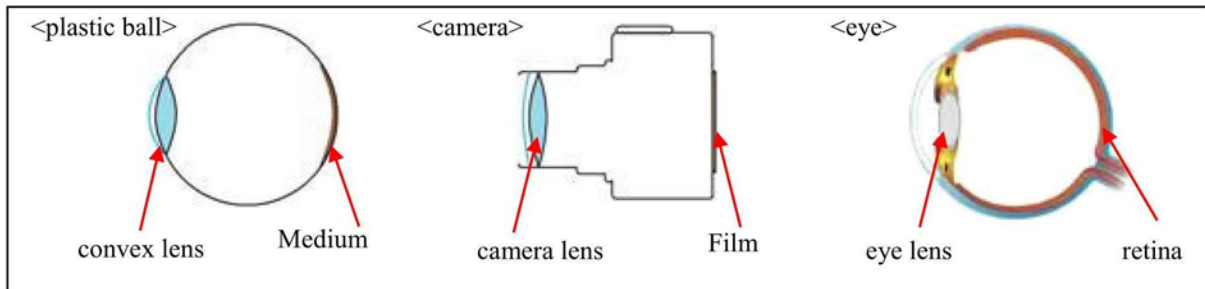


Because

- (a) After light passes through eye lens, it would converge inward and appear the image on the retina.
- (b) After light passes through eye lens, it would be reflected and the eye could see any image.
- (c) After light passes through eye lens, it would go straight forward and there would be no image on the retina.
- (d) After light passes through eye lens, it would spread outward and appear as the fuzzy image on the retina.

### Appendix 3. Concept Dependent Analogical Reasoning Test (CDART) Example Question

Of the following diagrams, which statement is correct regarding the similar function among plastic ball, camera, and eye?



- (a) Medium of plastic ball, film of camera, and retina has a similar function.
- (b) Medium of plastic ball, camera lens, and eye lens have a similar function.
- (c) Convex lens, camera lens, and retina have a similar function.
- (d) Convex lens, camera film, and eye lens have a similar function.

Because

- (a) Convex lens, camera lens, and retina all could show image.
- (b) Medium of plastic ball, camera lens, and eye lens all could refract light inward.
- (c) Medium of plastic ball, film of camera, and retina all could show image.
- (d) Convex lens, film of camera, and eye lens all could refract light inward.



## The Increasing Acceptance of Onscreen Marking – the ‘Tablet Computer’ Effect

David Coniam

Department of Curriculum and Instruction, Faculty of Education and Human Development, The Hong Kong Institute of Education, Hong Kong // coniam@ied.edu.hk

(Submitted June 20, 2012; Revised September 16, 2012; Accepted October 09, 2012)

### ABSTRACT

This study extends recent investigations of onscreen marking (OSM) in the Hong Kong public examination context. It examines whether marker attitudes, which have shown increasing acceptance of OSM since it was first introduced in 2007, continue to develop this acceptance in a different context. In 2012, all public examinations were marked onscreen. While previous studies investigated the attitudes of mature markers marking extended essay answers for single subjects (English and Liberal Studies), the current study examines the attitudes and perceptions of much younger markers when marking short, objective answer questions in the 2012 Hong Kong Advanced Level Examination Use of English. In addition, because the markers were younger than usual, the issue of whether their attitudes to OSM were affected by the ownership of a tablet computer was also investigated. The findings, which have implications for other examination jurisdictions, indicate a growing acceptance of OSM, especially by tablet owners.

### Keywords

Onscreen marking, Tablet computers, English language assessment

### Introduction

This study furthers the research conducted into OSM in the context of a jurisdiction where almost the entire public examination system is being marked on screen. The current study – investigating marker attitude towards OSM – extends previous studies of single examination subjects in the Hong Kong context (Coniam, 2009; Coniam and Yeung, 2010, for example), which showed an increasing but slow acceptance of OSM, in a number of ways. Coniam (2009) showed that markers of Grade 11 English Language examinations were consistently reliable whether marking on paper or on screen. Their attitudes to OSM revealed that they were prepared to ‘buy in’ to the new marking medium. Coniam and Yeung (2010), which examined the attitudes of markers of Liberal Studies, revealed similar reliability whether marking on paper or onscreen. Further, on all major indicators, the attitudes of Liberal Studies markers towards OSM were more positive than in the previous study for English.

First, previous studies examined markers who were marking (for English and Liberal Studies) *extended essays* on screen. The current study differs from previous studies in that it examines the marking of *short answer questions* in the 2012 Year 13 Hong Kong Advanced Level Examination (HKALE) Use of English examination. Short answer questions are targetted by Shaw (2008) as an area requiring further research when attempting to confirm the reliability of marking such questions onscreen compared to on-paper marking. The current study therefore contributes to our understanding of this area.

Second, previous studies focused on experienced markers who had all been marking on paper for a number of years. The subjects in the current study were younger – around 20 years of age – for the most part undergraduate language or linguistics students, the majority of whom had never marked before. The sample comprised the entire marking panel of 289 markers who all marked onscreen.

Third, the previous onscreen studies generally involved markers in marking for three or four hours a day over an extended period, e.g., two weeks. The current study involved one lengthier period of time working on screen. Specifically, the marking of the entire paper (40,000 candidates’ scripts) was completed *in a single day*. This required markers to work on screen for approximately six hours in that single day, each marker marking approximately 120 scripts, or 6,000 items.

Fourth, previous studies could not investigate the phenomenon of the ownership of tablet computers and whether such ownership affected markers' attitudes to OSM because ownership of such devices was not prevalent until 2010.

## Background to the study

The Hong Kong secondary school curriculum underwent significant restructuring in 2009 when the British 5+2 secondary system (with examinations at Years 11 and 13) was replaced by a 3+3 year system (similar to the Chinese and Australian education models), with secondary education lasting six years. The examination corollary was a single examination at the end of Year 12 (age 17-18) – the Hong Kong Diploma in Secondary Education (HKDSE) – the annual candidature for which is in the region of 80,000.

In addition to the structural changes to the education and examination system, a major operational change accompanying the introduction of the HKDSE, is that paper-based marking of public examinations ceases, with *all* public examinations marked on screen from 2012 onwards.

The Hong Kong Examinations and Assessment Authority (HKEAA), Hong Kong's public examination body, has, for some time, been investigating a variety of procedures and processes related to computerisation, including a dedicated onscreen marking (OSM) system. In late 2005, the Hong Kong SAR Legislative Council, (Legislative Council Panel on Education, 2005), allocated approximately US\$25 million towards the IT modernisation of the HKEAA.

The first step towards the adoption of OSM was taken in 2007 when all Year 11 Hong Kong Certificate of Education Examination (HKCEE) English Language and Chinese Language scripts were marked on screen. To investigate and validate the adoption of OSM, a series of studies were subsequently conducted with key subjects to compare the two modes of marking. In the first of these, Coniam (2009, 2011) reported quantitatively and qualitatively on marking comparability issues and marker attitude for the English language writing examination. He found that the reliability of marking was consistent between the two modes of marking and that, by 2011, OSM was becoming increasingly accepted.

The second subject to be investigated was Liberal Studies, a compulsory subject in the HKDSE, and a subject whose candidature has increased from its previous small size of 3,000 candidates to 80,000. Since the acceptance of Liberal Studies by parents and teachers was a priority for the HKEAA, with substantial public concern expressed over the introduction of this subject (Tsang, 2006), a number of different studies were consequently conducted for the subject. The acceptance by markers of OSM as the marking mode was one of the focuses (see Coniam & Yeung, 2010). Again, it was found that the reliability of marking was consistent between the two modes of marking.

The current study takes place in 2012, when all papers in all subjects are being marked on screen. The current study is, however, not simply a recycling of past studies. Timing is a key issue – given that 2012 is the milestone for OSM for all subjects. In part, the context for the current study focuses on the acceptance of the onscreen system – the issue of 'consumer validity' (Coniam, 1999), i.e., that markers are prepared to 'buy' the new onscreen marking system. The wholesale marking of papers onscreen and the findings of studies such as this have implications for other countries both in the region and elsewhere where OSM is either being practiced albeit on a smaller scale or being implemented.

The OSM marking process in Hong Kong is as follows. After each examination, scripts are delivered to a dedicated scanning centre where scripts are scanned and images saved. Markers then need go to special OSM centres (of which Hong Kong has seven at strategic locations) where they mark at dedicated workstations using a purpose-built system.

The OSM system requires markers to progress first through *training* and *qualifying* (demonstrating that their marking is to the appropriate standard), before finally moving to independent *live marking*. During live marking, 'control scripts' (with scores agreed by chief examiners) are regularly issued to markers so that their marking standard can be monitored. A further reliability-enhancing feature of the OSM system is that markers may view their own marking statistics (marking speed, mean score, mark distribution spread) and observe how their control script marking statistics compare with those of the chief examiners'.

For a discussion of onscreen marking with regards to procedures, benefits and drawbacks, the reader is also referred to Coniam (2009).

Another, external, factor is, however, driving the current study. This is the tablet computer, which became publicly available in 2010. This device has captured the public's imagination and, it is hypothesised, may play a role in the greater acceptance of OSM, providing substantially less resistance and negativity towards reading on screen. While, as reported, markers involved in the Hong Kong context over the past five years have been showing greater acceptance of OSM as a medium, they have continually commented on 'eye tiredness' as a feature of OSM that concerns them. Indeed, prior to the introduction of the computer tablet, onscreen reading has generally not had a particularly positive press; (Enright et al, 2000), for example, state that reading on screen is "generally less appealing than reading from paper" (p. 41). This observation has generally been borne out in research findings – at least those for reading in one's first language – because it appears that reading rates fall when onscreen reading is compared with traditional print material (Kurniawan and Zaphiris, 2001a; 2001b). To be fair, however, as Muter and Maurutto (1991) noted a while back, a multiplicity of factors account for why reading on screen is perceived to be more difficult than reading on paper.

A key factor in the current study, then, is the possession of a tablet computer, which, it will be suggested, is leading to a change in attitudes towards onscreen reading. Consequently, a brief examination of the introduction and concept of tablets, as well as very recently reported studies into their use as enhancers, or motivators of reading will now be provided. It is important to note that the markers did *not* mark the examination questions on a tablet. What is of significance is whether attitudes to marking onscreen were affected by ownership of a tablet computer.

### **Tablet computers – background and their takeup in the field of education**

Tablet computers have quite an extensive history, stretching back to the 1980s. Previously, tablets failed because of problems with weight, battery life, and a lack of software features such as finger and virtual keyboard support. In addition, they contained few applications specific to the platform.

With Apple's release of the first iPad in April 2010, tablets began to come of age and to be accepted on a massive scale by consumers. Indeed, estimates are that, in the two-year period April 2010-April 2012, some 67 million iPads were sold (<http://ipod.about.com/od/ipadmodelsandterms/f/ipad-sales-to-date.htm>).

The tablet computer has captured the public's imagination, and studies reporting its educational potential have already begun to appear – albeit limited in scope at this early stage in the device's life. Larson (2009) – a proponent of new digital literacies – reports on a project with students successfully incorporating e-books into an otherwise traditional literacy programme with Year 5 students in the U.S. (2009, pp. 256). McClanahan et al. (2012) report on the use of a tablet with an American Year 5 reader suffering from ADHD (Attention Deficit Hyperactivity Disorder). They report that the device not only helped the student to focus his attention, but that the student gained in confidence and in his sense of being in control of his learning. Further, pre- and post-tests indicated a one-year gain in reading in a six-week period.

Maynard (2010) reports on a study in the U.K. into the impact of e-books on young children's reading habits. In the study, 12 subjects – six children aged 7-12 and their parents – used an e-reader for a two-week period. In follow-up interviews, while all six adult participants reported still favouring traditional printed books, the children were split 50-50, with half preferring e-books over printed books. The group of children favouring e-books included one child who was a 'reluctant' reader and who subsequently reported substantial interest in reading.

Studies have, however, also reported mixed results. Sheppard (2011) reports on a study in Australia with 43 Year 6 boys reading e-books on iPads. From the crossed design study (where half read traditional printed books while half read on screen, and then vice versa), Sheppard reports that students participating in the study reported an increase in engagement when using a tablet. There was, however, not a corresponding rise in achievement, with comprehension scores actually decreasing with less able readers. Teachers also reported having "mixed feelings" about the use of the e-Book reader, with some commenting that at times it was simply a distraction.

Meurant (2010) discusses the use of tablets to promote student digital literacy in the context of English language learning in Korea – one of the most wired nations on the planet ([http://www.pbs.org/frontlineworld/stories/south\\_korea802/](http://www.pbs.org/frontlineworld/stories/south_korea802/)). He envisages the device helping teachers move from a traditional, static, use of the classroom to a more blended or hybrid form of education that combines traditional classroom instruction with computer-based language learning.

Early reviews of tablets were dismissive but such views would, however, appear to be a minority view – certainly from the perspective of the younger markers in the current study.

### **OSM – score comparability and marker attitude**

Two major issues – which have been reported on quite extensively in the acceptance of OSM – are score comparability and marker attitude (see, for example, Shaw, 2008; Coniam, 2009). The two issues will be briefly covered, although for a more in-depth discussion, the reader is referred to Coniam (2009).

While studies by Whetton & Newton (2002) and Royal-Dawson (2003) have pointed to certain differences between the two modes of marking, in general, little has been reported in the context of differences between onscreen and on-paper marking. In, for example, Powers et al.'s (1997) early research into OSM in the USA, no differences emerged between scores awarded in either medium. In the UK, studies by Newton et al. (2001), reported no difference in the overall ratings of either onscreen or on-paper groups. While Sturman & Kispal (2003) reported quantitative differences between onscreen and on-paper marking, they concluded that onscreen marking was as accurate as conventional marking. Bennett (2003) concluded that research results indicated minimal differences between the two mediums. Although Shaw (2008) reported minor differences between onscreen and on-paper marking, the differences were not significant. In general, results of more recent studies point towards the comparability of paper and screen-based marking in terms of the reliability of the results between the two modes of marking (Fowles, 2008; Johnson et al, 2010; Coniam, 2009).

Marker reactions to OSM were initially mixed. In Powers et al.'s (1997) study, the OSM system was received “relatively positively” by most markers. In contrast, Zhang et al. reported mixed reactions in their 2003 study, as did Fowles and Adams in their 2005 study. Twing et al. (2003) reported anxiety about marking on screen due to lack of computer proficiency among markers. Reactions towards OSM have, however, been improving with the passing of time, and there is a growing body of evidence to support the increasing acceptance of OSM by markers. In addition to studies by Coniam & Yeung (2010), the reader is also referred to Shaw (2008) and Shaw and Imam (2008).

### **The study**

This section describes the data which make up the study and the methods used to analyse the data.

The data for the study are drawn from the 2012 HKALE Use of English examination, where the candidature was 39,807. The paper under examination is Paper C (henceforth UEC), which is an objective reading / language systems paper. The paper has six parts, comprising, in 2012, 97 objective right/wrong items. The first three parts (consisting of 46 items) are multiple-choice, and are machine marked. The remaining three parts comprise short answers which need to be marked right or wrong, i.e., ‘1’ or ‘0’ entered in the appropriate box beside the item. The marking is therefore more ‘mechanical’ than marking extended essays, only requiring ‘expeditious reading strategies’ (Shaw, 2008, pp. 268). The English and Liberal Studies marking, described in previous studies, involved reading extended essays on screen, where candidates are awarded (in the case of English) scores from 1 to 6 in four domains. While essays need, to an extent, to be pondered over and the content considered (Shaw, 2008, pp. 268), rather less is required in the way of scoring short answers. Annotating scripts is not necessary; scores simply need to be entered for each of the four domains. The marking of short answers is, however, demanding in a different way. While the marking of the UEC paper requires less processing of extended texts than does the reading of long essays, considerable concentration and attention is nonetheless required if 51 items are to be processed and scored accurately over a period of six hours.

The majority of Hong Kong public examinations are marked by experienced teacher markers. Since the marking of the UEC paper is more mechanical than that required for other papers, the HKEAA engages a marking panel which comprises undergraduate language or linguistics major students from local Hong Kong universities. The UEC paper is marked in a single day, at a single marking centre, by approximately 300 markers. Markers are paid for a day's work – approximately six hours' worth of marking: from 9 a.m. to 4 p.m. with a break for lunch (and rest breaks as necessary) – so speed of marking per se is not an issue.

In the current study, the methodology adopted for the Liberal Studies study (Coniam and Yeung, 2010) is developed and extended. In the 2010 study (involving the marking of extended essays), all markers (N=46) of the 2009 HKALE Liberal Studies examination completed pre- and post-marking attitudinal questionnaires to gauge their perceived technological proficiency and attitudes towards OSM. Questions were posed on a 6-point Likert scale, with '1' indicating a positive response or agreement, and '6' a negative response or disagreement. With regard to computer proficiency – an important element in the adoption of a computerised marking system – markers generally rated themselves as competent, responding more positively than in the previous study with English language markers (Coniam, 2009). Attitudes towards OSM were positive, with markers more upbeat on certain key questions in their post-marking responses than in their pre-marking questionnaires. There were, however, certain issues about which markers were less positive. In particular, *eye tiredness through marking on screen* received a highly negative score of 4.95/6. Further, regarding preference for marking on screen or on paper, markers were strongly skewed towards the 'on paper' option. Nonetheless, the picture that emerged was that, overall, Liberal Studies markers were more positive towards OSM than were English language markers in the earlier study.

In the current study, the two-section format is maintained and adapted to the UEC environment, utilising both pre- and post-marking questionnaires. Where the pre-marking questionnaire diverges from the Liberal Studies one, however, is that a number of items ask markers whether they own a tablet computer, for how long, what they use it for and their attitude to onscreen reading since owning the tablet. A further issue investigated – which had been commented upon in interviews with Liberal Studies markers (Coniam, 2011) – was the extent to which markers felt under pressure because of constant monitoring and feedback from the computer system. Since in all previous studies, markers clearly view themselves as technically proficient (scrolling, enlarging, data entry etc), computer proficiency is generally taken as a given, and little attention is paid to the issue in this study. The focus is on markers' attitudinal responses, specifically, the eight items listed below:

- Rating of one's own computer proficiency
- Comfortable reading off screen
- Concerns about eye tiredness
- Break-taking frequency
- Expectations / evaluation of onscreen marking experience
- Attitude regarding the move from paper-based to onscreen marking
- Preference: marking on screen or on paper
- Feeling under pressure, due to being constantly monitored

## Research questions

As mentioned, the marker focus in the current study is different from previous studies in that subjects are younger markers who in the main have not previously marked for the HKEAA. In addition, these markers are marking for an extended period – six hours or more in a single day. The study pursues the hypothesis that younger markers will, as a starting point, be more predisposed to marking on screen than will older markers. The study is also hypothesising that with the massive take-up of tablet ownership, the younger generation of markers will be even more predisposed towards OSM, with ownership of a tablet resulting in a more positive attitude towards OSM. It should be restated at this juncture that it is ownership of a tablet that is being studied *not* marking onscreen on a tablet.

There are, consequently, three hypotheses in the current study which relate to marker attitude.

- Markers of the 2012 UEC examination will be more positive towards the OSM medium in their post-marking questionnaires than were HKALE Liberal Studies markers in the 2010 study.
- Markers will be more positive towards the OSM medium in their post-marking questionnaire than in their pre-marking questionnaire.

- Markers who own a tablet will be more positive towards the OSM marking medium in both their pre- and post-marking questionnaires than markers who do not own a tablet.

## Results and findings

The reliability (using Cronbach's alpha) for the attitudinal questions on the two questionnaires was first examined. A figure of 0.78 emerged for the 12 attitudinal questions for the pre-marking and a figure of 0.81 for the 14 post-marking attitudinal questions. Given that the level of 0.8 is generally suggested as desirable in a questionnaire (e.g., Dörnyei, 2003), this suggests that the questionnaires were generally well constructed.

In the analysis below, markers' background details are first presented, with a subsequent examination of the questionnaires. First, to provide a contrastive picture of whether attitudes are changing, comparisons are drawn between pre- and post-marking questionnaire results of the 2012 UEC markers against the 2010 Liberal Studies markers. Subsequent analyses of results focus on the pre- and post-marking results of markers who do or do not own a tablet.

### Markers, test takers and modules

324 markers were recruited by the HKEAA to mark the UEC paper. Ultimately 289 turned up for the whole day's work at PCs at the dedicated marking centre. The markers for UEC are university undergraduates studying, for the most part, towards a relevant language-related degree. Since completed pre- and post-marking questionnaires were retrieved from 280 markers, it is this figure (N=280) that forms the basis for discussion in the current study. The majority were marking for the first time, with 198 (71.1%) being first time markers. 67 (23.9%) had one year's OSM experience, while 14 (5.0%) had two years or more OSM experience.

Regarding age spread, 74 markers (26.4%) were below 20 years of age, 204 (72.9%) were aged 21-25, with only 2 (0.7%) aged 26 or more. The male-female ratio was comparable to the general English language teaching cohort in schools, with 221 (78.9%) respondents female.

Concerning tablet ownership, 82 of the 280 (29.3%) markers reported ownership.

Table 1 first presents comparative results of post-marking results by the UEC and Liberal Studies markers, with independent sample t-tests assuming equal variances (confirmed by Levene's test) as the statistic. It should be noted in Table 1 that, while the sample sizes are different, the samples constitute the entire marking panel for each subject. Item 8 does not appear in Table 1 since it did not form part of the questionnaire administered to the Liberal Studies markers.

It will be recalled that '1' generally indicated a positive response and '6' a negative response. 3.5 is therefore the mid-point of the 6-point scale. Consequently, a response mean of less than 3.5 can be taken as an indicator of positive attitude and greater than 3.5 an indicator of negative attitude.

*Table 1. UEC and Liberal Studies post-marking comparisons*

Question	Markers	N	Mean	SD	t-test results
1. Rating of one's own computer proficiency 1=high; 6=poor	UEC	279	2.17	0.89	n.s.
	LS	46	2.26	1.02	
2. Comfortable reading off screen 1=comfortable; 6=uncomfortable	UEC	279	2.56	0.97	t=-4.086, df=323, p=.000
	LS	46	3.22	1.22	
3. Concerns about eye tiredness	UEC	279	4.08	1.19	t=-4.644,

	LS	46	4.96	1.11	
4. Break-taking frequency 1=infrequently; 6=frequently	UEC	279	2.95	1.08	t=-7.575, df=323, p=.000
	LS	46	4.22	0.81	
5. Evaluation of onscreen marking experience 1=good; 6=bad	UEC	279	2.52	0.85	t=-4.464, df=322, p=.000
	LS	45	3.18	1.26	
6. Preference: marking on screen or on paper 1=on screen; 6=on paper	UEC	278	2.97	1.36	t=-6.829, df=322, p=.000
	LS	46	4.46	1.54	
7. Attitude regarding move from paper-based to onscreen marking 1=a good move; 6=a bad move	UEC	278	2.60	1.02	t=-5.278, df=322, p=.000
	LS	46	3.52	1.44	

As Table 1 indicates, with the exception of item 1 where both groups were equally confident of their computer proficiency, the UEC markers were significantly more positive than the Liberal Studies markers on every item. Furthermore, whereas with the Liberal Studies markers only two items indicated positive attitude, with the UEC markers, all items – with the exception of eye tiredness (item 3) – were on the positive side of the 3.5 mark. And even with this item, the response dropped from a strongly negative 4.96 to 4.08.

The two indicators of general acceptance – items 6 and 7 – show markedly positive shifts. Whereas Liberal Studies markers' response on item 7 of 3.52 indicated ambivalence, UEC markers' 2.60 showed an endorsement of the move to OSM. Regarding preference, UEC markers showed a clear preference for marking on screen with a mean of 2.97 whereas Liberal Studies markers were clearly in favour of paper marking at 4.46.

To illustrate the extent to which markers were satisfied by their marking experience, Table 2 presents the results of pre- and post-marking response means for the UEC markers, using paired t-tests. It is proposed that a positive shift will indicate greater 'consumer validity' – the validity of subjects' attitudes and feelings towards a test (see Coniam, 1999). An additional item included in Table 2 relates to how much pressure markers felt. Such a question was included because the OSM system reports on marker speed, reliability etc., with markers required to do additional retraining scripts to bring them back into line if their marking diverges from accepted standards.

*Table 2. UEC markers' pre- and post-marking attitudinal comparisons (N=280)*

Question	Phase	Mean	SD	t-test results
1. Rating of one's own computer proficiency 1=high; 6=poor	Pre	2.46	0.95	t=5.420, df=279, p=.000
	Post	2.14	0.88	
2. Comfortable reading off screen 1=comfortable; 6= uncomfortable	Pre	2.48	1.16	n.s.
	Post	2.55	0.97	
3. Concerns about eye tiredness 1=not concerned; 6=concerned	Pre	4.40	1.23	t=3.787, df=279, p=.000
	Post	4.08	1.19	
4. Break-taking frequency 1=infrequently; 6=frequently	Pre	3.43	0.87	t=6.974, df=279, p=.000
	Post	2.95	1.08	
5. Expectations / evaluation of onscreen marking experience 1=positive; 6=negative	Pre	2.79	0.83	t=4.420, df=279, p=.000
	Post	2.52	0.85	
6. Preference: marking on screen or on paper 1= on screen; 6=on paper	Pre	3.82	1.38	t=11.193, df=278, p=.000
	Post	2.96	1.36	
7. Attitude regarding move from paper-based to onscreen marking 1=a good move; 6=a bad move	Pre	2.86	0.98	t=4.726, df=277, p=.000
	Post	2.61	1.02	
8. Feeling under pressure, due to being constantly monitored 1=no pressure; 6= a lot of pressure	Pre	3.51	1.23	t=9.192, df=279, p=.000
	Post	2.90	1.18	

As can be seen from Table 2, UEC markers were more positive post-marking than they were pre-marking on all items, with the exception of how comfortable they were reading on screen (item 2) – although the difference was not significant. Eye tiredness was less of a concern, although it was still above the mid-point 3.5, indicating that there may still be an issue with staring at the screen for six hours a day. Preference for the OSM medium (item 6) shows a strong shift, moving from a negative-oriented 3.82 to a positive 2.96.

Markers' post-marking results also indicated that they had felt less pressure than they expected (item 8), revealing that they see the system helping them in their marking, rather than acting as a kind of 'big brother' check, hanging over them for the sole purpose of criticising them.

The sections so far have shown a contrastive picture of the acceptance of OSM. The following analyses will now explore the degree to which owning a tablet affects attitudes. Pre- and post-marking responses are given separate treatment. Table 3 first presents the picture of pre-marking attitudes in the light of whether markers own a tablet or not, using independent sample t-tests as the statistic.

*Table 3. UEC markers' pre-marking attitudes, pad computer owners vs. no pad*

Question	Own pad?	N	Mean	SD	t-test results
1. Rating of one's own computer proficiency 1=high; 6=poor	Yes	82	2.24	0.95	t=-2.440, df=278, p=.015
	No	198	2.55	0.91	
2. Comfortable reading off screen 1=comfortable; 6= uncomfortable	Yes	82	2.32	1.13	n.s.
	No	198	2.55	1.16	
3. Concerns about eye tiredness 1=not concerned; 6=concerned	Yes	82	4.28	1.32	n.s.
	No	198	4.45	1.19	
4. Break-taking frequency 1=infrequently; 6=frequently	Yes	82	3.28	0.91	t=-1.853, df=278, p=.065
	No	198	3.49	0.85	
5. Evaluation of onscreen marking experience 1=positive; 6=negative	Yes	82	2.79	0.94	n.s.
	No	198	2.79	0.78	
6. Preference: marking on screen or on paper 1= on screen; 6=on paper	Yes	82	3.54	1.46	t=-2.279, df=278, p=.023
	No	198	3.94	1.33	
7. Attitude regarding move from paper-based to onscreen marking 1=a good move; 6=a bad move	Yes	82	2.73	1.10	n.s.
	No	197	2.93	0.93	
8. Feeling under pressure, due to being constantly monitored 1=no pressure; 6= a lot of pressure	Yes	82	3.39	1.26	n.s.
	No	198	3.56	1.11	

While five out of eight differences between tablet-owning and non-tablet-owning markers are not statistically significant, the mean scores reveal a uniformly clear trend toward positive attitudes. As can be seen from Table 3, on every item, markers who own a tablet are more positively disposed to OSM than markers who do not. Markers who own a tablet see themselves as being more proficient with computers (item 1) and having a significantly stronger preference for marking on screen – 3.54 as against 3.94 (item 6). Marginal significance also emerges for concern about eye tiredness (item 3) – although this is still in negative territory with both groups.

Finally, Table 4 presents the picture of post-marking attitudes in the light of tablet ownership, again with independent sample t-tests as the statistic.

*Table 4. UEC markers' post-marking attitudes, pad computer owners vs. no pad*

Question	Own pad?	N	Mean	SD	t-test results
1. Rating of one's own computer proficiency 1=high; 6=poor	Yes	82	2.04	1.02	n.s.
	No	198	2.18	0.82	
2. Comfortable reading off screen 1=comfortable; 6= uncomfortable	Yes	82	2.33	1.01	t=-2.491, df=278, p=.013
	No	198	2.65	0.94	
3. Concerns about eye tiredness 1=not concerned; 6=concerned	Yes	82	3.98	1.15	n.s.
	No	198	4.12	1.21	
4. Break-taking frequency 1=infrequently; 6=frequently	Yes	82	2.95	1.10	n.s.
	No	198	2.95	1.07	
5. Evaluation of onscreen marking experience 1=positive; 6=negative	Yes	82	2.45	0.88	n.s.
	No	198	2.55	0.84	
6. Preference: marking on screen or on paper 1= on screen; 6=on paper	Yes	81	2.96	1.46	n.s.
	No	198	2.97	1.32	



7. Attitude regarding move from paper-based to onscreen marking 1=a good move; 6=a bad move	Yes	81	2.43	1.05	t=-2.047, df=277, p=.042
	No	198	2.70	1.00	
8. Feeling under pressure, due to being constantly monitored 1=no pressure; 6= a lot of pressure	Yes	82	2.80	1.21	n.s.
	No	198	2.94	1.19	

While significance only appears on two items, again, on all items, tablet-owning markers are more positive (with the exception of item 4, the need to take a break). Tablet owners are significantly more disposed to reading on screen (2.33) than those who do not own them (2.65). They are also significantly more positive about the move to OSM. These findings echo the findings of studies involving children reading on tablets earlier in this article. The results in Table 4, while showing cautious support for the fact that owning a pad is going to make a marker more positive, still reports very encouraging results for both groups' attitudes towards OSM. This may well be the youth factor. As indicated earlier, all the markers are confident about their computer proficiency. This is not surprising as they have been accustomed to using PCs and laptops for well over a decade. Furthermore, simple observation of behaviour in university student common rooms and cafeterias indicates that tablet owners are reading off screen a great deal. Much more onscreen reading is taking place every single day than hitherto. It is not the same as the utterly novel activity encountered by experienced markers who had never marked onscreen before in the 2009 and 2010 studies.

## Conclusions

This study has reported on the marking of Section C of the HKALE Use of English examination, a paper comprising 51 short answers which is marked in one day by approximately 300 undergraduate students from local Hong Kong universities.

The study was pursuing three hypotheses. The first was that with the passing of time, markers would show greater acceptance of the OSM system. Specifically, it was hypothesised that the 2012 UEC markers would exhibit a more positive attitude towards OSM than the 2010 Liberal Studies markers. With regard to this hypothesis, the 2012 UEC markers were more positive towards OSM than the 2010 Liberal Studies markers on virtually every count. Their responses are clearly skewed towards the more positive end of the spectrum – i.e., many items recorded a mean on the positive side of 3.5. This hypothesis was accepted.

The second hypothesis, that the 2012 UEC markers would be more positive after their marking experience than before it (i.e., in their post-marking questionnaires than in their pre-making ones) was also accepted. UEC markers were more positive post-marking on all items bar one – comfort with reading on screen.

The third hypothesis was that with the recent intense interest in tablets, and the manner in which tablets have taken the world by storm – especially for the younger generation – young markers who owned a tablet would be more disposed towards OSM as a medium than would markers who did not own one. Tablet owners were more positive on almost every item, both pre- and post-marking. This hypothesis was also accepted. The implications of this are clearly that ownership and regular use of tablet computers is predisposing users towards reading on screen, and consequently appears to have had a knock-on effect of greater predisposition towards acceptance of onscreen marking. This finding is significant for those countries who are implementing OSM. Further investigations in other examination jurisdictions should be launched to see if the results of this study are generalisable. If they are, it is likely that in a relatively short time frame, OSM will become universally accepted around the world.

It was stated that while the two modes of marking are different, they put different demands upon markers. Marking extended essays requires markers to process lengthy pieces of text; their marking requirements are usually limited, however, to a small number (four or so) of scores to be awarded. In contrast, the UEC markers had to mark 51 objective items, where, while less judgment was called for than with the reading of extended texts, each item required a score to be input. Consequently, marking objective items is also demanding, albeit in a different manner.

A final issue concerns a virtual truism – the issue of age, long recognised as being a factor in the acceptance of computers. In Czaja and Sharit's (1998) study, for example, older people reported less comfort and less competence with computers, as well as perceiving computers as being more 'dehumanising' than their younger counterparts. Similarly, age has emerged as a factor in previous OSM studies. In the (Coniam, 2010) study with established Liberal Studies teacher markers, for instance, younger markers were more positive concerning the introduction of OSM than were older markers. And in the current study, UEC markers were considerably younger than the whole cohort of Liberal Studies markers.

Given that the younger markers will become teacher markers of other papers including essay-type answers and extended response answers with the passing of time, their upbeat attitude towards the system bodes well for eventual broad acceptance of OSM both in Hong Kong and in other countries. An area for further research could involve investigating whether older, more experienced markers who possess tablets are more predisposed to OSM than those who do not own a tablet.

As OSM is being introduced for the majority of subjects in 2012 with the implementation of the HKDSE, it is therefore vital that the system be accepted by markers. The results of the current study would suggest that this is likely to be the case. It further suggests that if these findings can be replicated in other countries and regions, the outlook for general acceptance of OSM is encouraging. In addition to the findings of previous studies, other countries can use the new conditions investigated here (e.g., younger markers, short answer responses to examination questions and possession of a tablet computer) to see if similar findings prevail in their jurisdictions. Above all, researchers in other countries can make use of the growing evidence from Hong Kong that, over time, markers become more predisposed to marking onscreen.

## Acknowledgements

I would like to thank the Hong Kong Examinations and Assessment Authority—in particular Christina Lee, the General Manager for Assessment Development, and Jennifer Shiu, the Manager for Assessment Development (English)—for support on the project regarding access to markers and for data collection.

## References

- Bennett, R. E. (2003) On-line assessment and the comparability of score meaning (ETS Report RM-03-05). Princeton, NJ: Educational Testing Service.
- Coniam, D. (1999). Subjects' reactions to computer-based tests. *Journal of Educational Technology Systems*, 27(3), 195-206.
- Coniam, D. (2009). A comparison of onscreen and paper-based marking in the Hong Kong public examination system. *Educational Research and Evaluation*, 15(3), 243-263.
- Coniam, D. (2011). A qualitative examination of the attitudes of liberal studies markers towards onscreen marking. *British Journal of Educational Technology*, 42(6), 1042-1054.
- Coniam, D., & Yeung, A. (2010). Markers' perceptions regarding the onscreen marking of liberal studies in the Hong Kong public examination system. *Asia Pacific Journal of Education*, 30(3), 249-271.
- Czaja, S. J., & Sharit, J. (1998). Age differences in attitudes toward computers. *Journal of Gerontology*, 53(5), 329-340.
- Dörnyei, Z. (2003). *Questionnaires in second language research: construction, administration, and processing*. Mahwah, NJ: Lawrence Erlbaum.
- Enright, M.K., Grabe, W., Koda, K., Mosenthal, P., Mulcahy-Ernt, P., & Schedl, M. (2000). *TOEFL 2000 reading framework: A working paper*. Princeton, NJ: Educational Testing Service.
- Fowles, D. (2008, September). *Does marking images of essays on screen retain marker confidence and reliability?* Paper presented at the International Association for Educational Assessment Annual Conference, Cambridge, UK.
- Fowles, D., & Adams, C. (2005, September). *How does assessment differ when e-marking replaces paper-based marking?* Paper presented at the 31st International Association for Educational Assessment Conference, Abuja, Nigeria.

- Johnson, M., Nádas, R., & Bell, J. F. (2010). Marking essays on screen: An investigation into the reliability of marking extended subjective texts. *British Journal of Educational Technology*, 41(5), 814-826.
- Kurniawan, S.H., & Zaphiris, P. (2001a). Reading online or on paper: which is faster? In C. Stephanidis (Eds.), *Proceedings of the Ninth International Conference on Human-Computer Interaction* (pp. 220-222). Mahwah, NJ: Lawrence Erlbaum.
- Kurniawan, S.H., & Zaphiris, P. (2001b). Investigating computer screen and paper reading speed differences. In W. A. Lawrence-Fowler & J. Hasebrook (Ed.), *Proceedings of WebNet 2001* (pp. 732-737). AACE Press: Charlottesville, VA.
- Larson, L.C. (2009). Digital literacies. *Journal of Adolescent & Adult Literacy*, 53(3), 255-258.
- Legislative Council Panel on Education. (2005). *Grant to support the modernization and development of the Hong Kong examinations and assessment authority's examination systems* (LC Paper No. CB(2)323/05-06(01)). Retrieved May 20, 2008, from <http://www.legco.gov.hk/yr05-06/english/panels/ed/papers/ed1114cb2-323-1e.pdf>
- Maynard, S. (2010). The impact of e-books on young children's reading habits. *Publishing Research Quarterly*, 26 (4), 236-248.
- McClanahan, B., Williams, K., Kennedy, E., & Tate, S. (2012). A breakthrough for Josh: How use of an iPad facilitated reading improvement. *TechTrends*, 56(3), 20-28.
- Meurant, R. C. (2010). The iPad and EFL digital literacy. *Communications in Computer and Information Science*, 123, 224-234.
- Muter, P., & Maurutto, P. (1991). Reading and skimming from computer screens and books: The paperless office revisited? *Behaviour & Information Technology*, 10(4), 257-266.
- Newton, P., Whetton, C., Adams, E., Bradshaw, J., & Wong, C. (2001). *An evaluation of the 2001 new technologies pilot*. Slough, UK: NFER.
- Powers, D., Kubota, M., Bentley, J., Farnum, M., Swartz, R., & Willard, A. (1997). *A pilot test of on-line essay scoring* (ETS Report RM-97-07). Princeton, NJ: Educational Testing Service.
- Royal-Dawson, L. (2003). *Electronic marking with ETS software. AQA Research Committee paper RC/219*. Manchester, UK: Assessment and Qualifications Alliance.
- Shaw, S., & Imam, H. (2008, September). *On-screen essay marking reliability: towards an understanding of marker assessment behaviour*. Paper presented at the IAEA 2008 Conference, Cambridge, UK.
- Shaw, S. (2008). Essay marking on-screen: Implications for assessment validity. *E-Learning*, 5(3), 256-274.
- Sheppard, D. (2011). Reading with iPads – the difference makes a difference. *Education Today*, 13(3), 12-15.
- Sturman, L., & Kispal, A. (2003). *To e or not to e? A comparison of electronic marking and paper-based marking*. Paper presented at the Annual International Association for Educational Assessment Conference, Manchester, UK.
- Tsang, W.K. (2006). *In search of the meanings for liberal studies in Hong Kong senior secondary education*. (in Chinese). Hong Kong: Hong Kong Institute of Educational Research, Faculty of Education.
- Twing, J., Nichols, P., & Harrison, I. (2003, June). *The comparability of paper-based and image-based marking of a high-stakes, large-scale writing assessment*. Paper presented at the IAEA Conference, Manchester, UK.
- Whetton, C., & Newton, P. (2002, September). *An evaluation of on-line marking*. Paper presented at the 28th International Association for Educational Assessment Conference, Hong Kong.
- Zhang, Y., Powers, D., Wright, W., & Morgan, R. (2003). *Applying the online scoring network (OSN) to advanced program placement program (AP) Tests*. Princeton, NJ: Educational Testing Service.

## Blog-Assisted Learning in the ESL Writing Classroom: A Phenomenological Analysis

Ming Huei Lin<sup>1\*</sup>, Nicholas Groom<sup>2</sup> and Chin-Ying Lin<sup>3</sup>

<sup>1</sup>Department of English, Tamkang University, Taiwan // <sup>2</sup>Center for English Language Studies, University of Birmingham, UK // <sup>3</sup>Department of Applied Foreign Languages, Chaoyang University of Technology, Taiwan // johnlin@mail.tku.edu.tw // n.w.groom@bham.ac.uk // cylin@cyut.edu.tw

\*Corresponding author

(Submitted May 26, 2012; Revised August 08, 2012; Accepted October 08, 2012)

### ABSTRACT

This study explores the experiences of a group of Taiwanese ESL student writers on a course of study that used a blog-assisted language learning (BALL) methodology. Data from in-depth interviews with eight participants were studied qualitatively using a phenomenological method of analysis. Participants' blogging activities were found to be inhibited by low L2 proficiency levels, by feelings of anxiety and embarrassment about possible peer reactions to their work, and by the slow pace at which completed compulsory blogging tasks were completed. It is argued that these findings help to explain the observation made in a number of previous studies, that second language writing students are enthusiastic about BALL in principle, but are not motivated to engage voluntarily in second language blogging activities in practice. The paper concludes by offering suggestions for ameliorating the problems identified in our analysis, and by identifying avenues for further research.

### Keywords

Classroom blogging, Blog-assisted language learning, ESL writing, Phenomenological analysis

### Introduction

Advances in information and communications technologies are now widely regarded as having huge potential for enhancing the teaching and learning of second language writing (Hyland, 2003; Warschauer, 2010). In recent years, the focus has begun to shift to Web 2.0 tools in general, and to the pedagogic use of blogs in particular (Campbell, 2003; Ward, 2004; Arani, 2005; Pinkman, 2005; Fellner & Apple, 2006; Warlick, 2007; Richardson, 2009; Wang, 2009; Arslan & Şahin-Kızıl, 2010; Sayed 2010; Sun, 2010; Lin, Lin, & Hsu, 2011; Lin, 2012; Hunter, Muilenburg, & Burnside, 2012; Nguyen, 2012; Sun, & Chang, 2012). In the field of second language teaching and learning, this interest has led to the development of a new field of research and practice, which Ward (2004) has termed *blog-assisted language learning* (henceforth BALL). The term BALL is used in different ways by different researchers and covers a wide variety of implementations in terms of software and course design considerations (cf. Ward, 2004; Cheng, 2006; Sun, 2010; Lin, 2012). For the purposes of this paper, however, BALL will be understood in general terms as covering any teaching and learning activities that involve the use of blogs as a computer-mediated platform (1) where interactions both within and beyond the classroom take place between teachers, students, and even (in principle, at least) members of the general public, and (2) where language learning activities are observable.

Although a comparatively recent phenomenon, BALL has already generated considerable excitement among second language writing researchers, teachers and course designers. BALL is seen as holding particular promise for developing learner autonomy, opening up communication pathways between teachers and students outside the classroom, and promoting motivation among students learning to write in a second language (Ward, 2004; Fellner & Apple, 2006; Iida, 2009; Noytim, 2010; Kang, Bonk, & Kim, 2011; Nguyen, 2012; Sun, & Chang, 2012; Taki & Fardafshari, 2012; Yunus, Salehi, & Chenzi, 2012). Some advocates of BALL have also suggested that these various advantages may be mutually reinforcing. That is, it has been proposed that students' motivation levels with regard to second language writing activities may be boosted by BALL precisely because it uses an exciting form of new technology that provides them with enhanced opportunities to engage in communication with teachers and peers (Pinkman, 2005; Kang, Bonk, & Kim, 2011; Trajtemberg & Yiakoumetti, 2011). These higher levels of motivation may in turn encourage students to engage more frequently - and more voluntarily - in BALL activities, thereby leading the students' learning into a virtuous cycle of the kind envisaged in contemporary motivation theory (e.g., Dörnyei, 2001; Dörnyei & Ushioda, 2011).

Though plausible enough in themselves, these optimistic conjectures are not entirely borne out by the findings of the empirical research that has actually been carried out on BALL to date. While studies have consistently found that student writers do indeed seem to hold positive beliefs in and attitudes toward BALL (Arani, 2005; Pinkman, 2005; Cheng, 2006; Sun, 2010; Trajtemberg & Yiakoumetti, 2011; Lin, 2012; Nguyen, 2012), it is equally clear from recent research that this enthusiasm does not seem to translate into any enhanced desire on the part of students to engage in blog-assisted writing activities (Chiao, 2006; Wu, 2008; Lin, Lin, & Hsu, 2011). In other words, students say that they like BALL, but they do not seem to be motivated to engage in any second language blogging activities unless they are obliged to do so by their teachers (cf. Wu, 2008; Miyazoe & Anderson, 2010). Even starker illustrations of this lack of take-up can be found in the student blogs conducted in recent studies by Arslan and Şahin-Kızıl (2010), Trajtemberg and Yiakoumetti (2011), and Lin (2012), which are still live and can thus still be visited on the Internet. From our own perusal of these websites at the time of writing of this paper (August 2012), it is clear that the students in these studies did not make any further posts to their own or their peers' blogs once the research projects motivating them had finished.

A key question that arises at this point, then, concerns *why* there is a disjunction between what students say when invited to evaluate BALL on the one hand, and what they actually do (or more precisely, do not do) when engaged in this new mode of learning on the other. Why does enthusiasm for BALL in principle fail to translate into blogging as self-directed language learning activity in practice? In the remainder of this paper we will report on an in-depth qualitative study that aims to cast light on precisely this question. For this purpose, a leading group of subsidiary research questions will need to be answered. These questions are as follows:

- What do the participants think about learning by using blogs?
- How do the participants describe their learning attitude and motivation when learning writing through using blogs?
- How do the participants evaluate the effectiveness of the blogging activities used in these writing lessons?
- Are the participants willing to attend more lessons supported with BALL like this? Why?/Why not?

## Methodology

### Research setting and participants

The present study was conducted in the Department of Applied Foreign Languages at a university in central Taiwan. Twenty-five first-year English majors agreed to do a BALL course by signing a consent form. The class met for two hours each week on Friday mornings, and the course aimed at enabling students to develop well-organized texts in different types of writing. The curriculum focused on details of language use, mechanics, content and organization, and related these to the three key student genres of narrative writing, comparison-contrast essays, and argumentative essays. Given that the aim of the research was to probe students' learning experiences with BALL in detail, the course was taught over a full academic year (i.e., two semesters, which amounts to 36 weeks in total in the Taiwanese educational system). Students were required to post at least 17 journal or assignment entries on their blogs in order to ensure an adequate degree of immersion in this experience.

### Course design

All classes were conducted in a computer laboratory where each individual student was provided with a computer connected to the Internet to facilitate in-class blogging activities. All of the participants had either a personal computer or a laptop at home, so after-class blogging activities were also feasible. Two types of blogs were concurrently used in the course: a tutor blog and learner blogs (Campbell, 2003). The tutor blog, created and maintained by the instructor, was used to deliver course information and material and share students' texts. The instructor also used the tutor blog to interact with and take questions from students both in and after class. The learner blogs were owned and maintained by individual student bloggers. Students were required to post their writing assignments and journals in their own learner blogs; they were asked to provide feedback on fellow students' entries; and the participants were encouraged to invite their friends to comment on their blog entries. The research instructor also marked and commented on all students' work online. As well as providing a complete record of all teaching input, the tutor blog also provides access (via its left sidebar) to the learner blogs and to all teacher feedback on students' work.

In the middle of the course, students were also asked to view an online writing lesson produced by the class instructor on how to write comparison-contrast essays. This lesson used both PowerPoint and video-taping and was embedded in the tutor blog. The students were encouraged to view and comment on this particular blogging lesson outside of class and in their own time using a computer connected to the Internet. Immediately after the end of course, eight students (six females and two males) volunteered to be interviewed about their experiences as learners.

## Data collection

The data collected for this study were comprised of in-depth interviews with eight student volunteers. The interviews took place on the same day and at the same location (the teacher's office), and the recruited participants were interviewed in turn. The interview duration for each participant was anticipated to be around 30 minutes. To facilitate the collection of data relating to the given aim, a group of open-ended questions corresponding to the research questions listed above was used.

It should be noted that, because the participants' English proficiency levels were relatively low (i.e., elementary to intermediate), all interviews were conducted in Mandarin Chinese. This ensured that all students were able to express themselves equally clearly, precisely and explicitly. (The interview extracts reported later in this paper will be presented in English translation, however.) In addition to the interviews, the interviewees' final journal entries were also collected as these related to the students' reflections on the course of their learning. Both of these data sources will be discussed in the analysis result.

## Data analysis

The particular qualitative approach used in the research reported in this paper is phenomenological analysis (Moustakas, 1994; Giorgi & Giorgi, 2008). The classic statement of this methodology is provided by Moustakas (1994). According to Moustakas, the analyst must begin by producing an exhaustive list of every interviewee statement that may be relevant to the experience under consideration. The analyst must then reduce and eliminate unclear statements from this list and sort the remainder into clusters of *invariant constituents*, statements that express single concepts or ideas. In phenomenological analysis, the unifying concepts that bind particular sets of invariant constituents together are known as *thematic portrayals*. These thematic portrayals in turn form the basis of a further distillation of participants' experiences into *textural* and *structural descriptions*. Textural descriptions focus on participants' thoughts and feelings with regard to these experiences, while structural descriptions attempt to account for the particular ways in which these feelings and thoughts are connected with the target experience. These two layers of interpretation are then combined in order to develop a final account of "the meanings and essences of the experience" for each participant in the research (Moustakas, 1994, p. 121). These individual textural-structural descriptions may then also be further synthesized into a *composite* textural-structural description, which presents the experience of the research participants as a group (McNamara, 2005).

## Results

This section will first report on the interview length for each participant. For reasons of space, however, this section will then present only the main analysis result by providing the composite textural-structural description of students' blogging experience. For relatively more accessible readability, appropriate subheadings will be used in the presentation of the composite textural-structural description.

### Interview length

An overall interview duration and relevant information is presented in Table 1 (pseudonyms are given for each student). It should be noted, however, that the interview length on average (around 26 minutes) was slightly shorter than the anticipated (30 minutes). This is particularly obvious with Zhang and Cindy, who gave relatively less detailed accounts during the interview. Although their cases may seem less suitable for phenomenological analysis, where lengthy transcripts are generally preferred, we decided to retain them here because the majority of their

thoughts and feelings about the experience were found to correspond to a very high degree with those of the other interviewees.

*Table 1. Interview Duration for Each Interviewee*

Names	Amy	Gloria	Zhang	Cindy	Olivia	Ginger	Danny	Isaac
Gender	Female	Female	Female	Female	Female	Female	Male	Male
Interview time (min/sec)	23'27	25'22	21'39	17'38	27'49	27'15	33'53	32'15
English proficiency level*	I	E	I	I	E	I	I	E
No. of journal entries used	2	2	2	2	2	2	2	2

\*Note. E stands for elementary; I stands for Intermediate.

## **Composite textural-structural description of the BALL group's experience**

### *Blogging as journaling*

The whole process of learning writing using BALL was often perceived as a means of learning writing by journaling. Interviewees frequently mentioned how they thought about doing more journaling in order to improve their writing, or how they felt that keeping a journal was a good way of practicing writing. This was also observed when students talked about their inactive learning patterns in terms of infrequent journaling. Interestingly, the fact that they were being required to post their journal entries to the Internet seemed to be less salient here; when reflecting on journaling as a learning activity, the students tended to focus less on the technological means through which the activity was being mediated, and more on why they were being required to do it.

### *Blogging as a fresh, novel, and convenient platform*

This is not to say that the students were indifferent to the innovative, computer-mediated format of the program, however. On the contrary, they described the experience of learning writing by blogging as a fresh and novel one. The integration of new technology into the learning setting left them perceiving the experience differently from the conventional, non-online approaches that they had experienced previously in their school careers. For example, Gloria reflected in her final journal entry that "this experience is so new and cool to me," a sentiment with which Cindy also agreed. To many others, BALL was perceived as novel because their entries were published in an online public setting.

This feature of BALL as a public online learning setting was also deemed convenient, and was one of the very first feelings that the participants responded to in relation to describing their learning experience. Isaac liked the fact that he could "take the advantage of retrieving past work", while Danny regarded the ability to access peer journals as "convenient." The underlying component of this sense of convenience, however, mostly related to the fact that they were able to submit assignments online, from any location, and in their own time. As Cindy remarked, "this is very convenient because I don't need to worry about forgetting to bring the material or rushing to school to hand in assignments to the office." A further perceived advantage was pointed out by Olivia: "blogs are convenient because I can access material or review your [referring to the teacher] notes any time I like." It is largely because of this general sense of convenience that students expressed approval at the idea of BALL being used in future courses.

### *Improved writing skills through blogging as a form of online mimicry learning*

The BALL experience also helped students to cultivate the belief that learning writing skills using blogs can indeed help them to write more effectively in English. This was seen primarily as a matter of noticing and/or appropriating 'the good bits' from peers' blog entries. "By reading other classmates' entries, I can review some English words that I have learned before," stated Amy; "[blogging] their entries also allows me to pick up some [ideas about] language usage." Similarly, Cindy pointed out how "I learned some English phrases for some [Chinese] expressions that I wanted to use but didn't know how to," and Ginger was glad that, by blogging others' journals, she had picked up some new words that would be helpful for her future writing. Students also said that they could brainstorm writing

ideas for their own assignments by reading peer work on blogs. According to Olivia, “removing this setting would lose me another way of [learning writing].”

*Increased awareness of limited linguistic ability woven with concerns about time and enlarged readership through blogging, in turn causing limited blogging patterns*

Throughout the BALL course, students were constantly reminded of their limited language ability and were often trapped in struggles over deciding whether to blog or not. These anxieties often went together with time management issues. Although one of the best student writers in the class, Ginger frequently abandoned journal entries before finishing them because of her concerns about the amount of time needed to ensure correct language use. Isaac and Cindy also found attending to the accuracy of their blog posts “very time-consuming and troublesome.” As seen earlier, Danny felt that he was “too busy” to post journal entries in English. Isaac did not even revise his entries at all according to the teacher’s feedback, as he felt that “revising an entry in English takes ages.”

The students’ consciousness of their low English proficiency levels was also clearly observed through their limited online communication model, which mainly took place through the *comment* function in the blogging software used in the class. Danny confessed that he seldom commented on peer entries because this was required to be done in English. Similarly, Cindy reported that, most of the time, she had “no idea about what to feed back because it has to be done in English.” Worried about her use of grammar, Ginger read peer entries without commenting as freely as she wanted to, restricting herself to anodyne messages of encouragement. Similar anxieties caused Amy to keep her supplementary journal entries offline, and to impose on herself a policy of “only posting grammatically correct messages.”

Students felt particularly embarrassed and worried about making grammatical mistakes. Specifically, they often found themselves feeling worried about readers noticing their errors and laughing at them. Zhang honestly admitted that she “would mock others’ mistakes, thinking how silly that writer is,” and felt that, in turn, her peers would be equally critical of her if they observed any mistakes in her blog entries. For similar reasons, Amy refused to share her learner blog with friends other than her classmates in this research. She disliked the idea that “there would be more people spotting my grammar mistakes ... They may laugh at me.”

The students’ awareness of a large and potentially critical audience for their work and their anxiety about grammatical mistakes did not affect the amount of effort they put into their writing, however. On the contrary, the students felt that their motivation levels were unaffected by the BALL method. As Zhang put it: “I am concerned that others may laugh at me ... but I think I write with the same attitude toward writing.” Essentially the same point was made by Isaac (“My attitude is the same ... when I am working on a piece of writing for blogs”) and Danny (“I wouldn’t deliberately write better because there are more readers ... I write what I would have written in any case”). Amy too reported that “I always write carefully and revise before submitting a piece of writing” and thus “did not perceive any difference” in her attitude toward writing using a learner blog.

*BALL as a toy with amusing colours (convenience, novelty, and freshness) for learning on the outside, but, unfortunately, the essence inside being perceived to be bland*

Another aspect of the BALL experience that did not seem to have much positive impact on the students’ learning attitudes and practices was their perception, discussed earlier, of BALL as a fresh, novel and convenient classroom approach. Although the students thought the BALL format was “cool”, it did not kindle in them any desire to engage in blogging activities unless they were required to do so. As both Gloria and Isaac remarked, “the convenience [of BALL] is nothing but convenience.” Even Danny, the most enthusiastic blogger in the group, saw blogging more as a new technological platform for reading and writing activities than as a new form of literacy practice in its own right.

*BALL serving as an informal but ineffective physical setting for learning*

Danny did, however, feel that he was writing for his own online space rather than for an assignment while engaged in BALL activities. He also commented that the teacher’s feedback on his blog felt more like the informal kinds of



electronic message that he received in the normal course of his daily life (e.g., via SMS messaging or Facebook) than the formal kinds of instructional feedback received in the traditional writing classroom.

This feeling of informality and casualness was also reported by other members of the interview group. To some extent, this aspect of the BALL approach was described in positive terms. In particular, interviewees felt that blogging helped them to perceive writing as a less dry and formal academic activity than hitherto. However, they also found the physical classroom setting (see Figure 1), in which each student was allocated one computer, unduly distracting. Ginger even reported that BALL left her feeling less incentivized compared to other methods of teaching, on the grounds that the classroom setting decreased student-teacher interaction: “The computer monitors block people’s sight, so you [the teacher] can’t see our faces, and this reduces the interaction [between the teacher and the class] ... it’s like ... I can only interact with people sitting next to me”.



*Figure 1: The classroom setting for the BALL study*

The rest of the interviewed participants admitted that they would spend class time checking email or doing online window shopping, instead of collaborating with each other on an in-class task or conducting blogging activities, as they were supposed to be doing. Even when the teacher was speaking to the whole class, many participants had a feeling that “the computer/Internet was distracting.” Zhang stated that she didn’t “really concentrate in class unless [the teacher] monitored our computers”. The distraction became even worse when the class was taught via the online lecture designed for them to view outside the classroom. All of the interviewees reported that they did this at home. Unfortunately, the majority found it difficult to concentrate on the online teaching input: “I skipped the lesson”; “I skimmed to the end for the assignment”; “I didn’t stay at my desk all the time”; “I just left the PowerPoint running and cleaned my room at the same time”. In other words, the students found that, whereas they were distracted *by* the computer when doing BALL activities in class, they were more likely to be distracted *from* the computer when doing BALL activities outside class.

## **Discussion and conclusions**

The phenomenological account of participants’ experiences of BALL presented above reveals a complex and at times seemingly contradictory set of meanings and essences that cannot easily be boiled down into a single overall positive or negative evaluation. On the one hand, the participants described BALL in very positive terms as a novel, convenient and refreshingly informal alternative to the traditional ESL writing classroom, and were enthusiastic about the idea of incorporating the BALL format into future classes. On the other hand, and at the same time, the participants expressed generally negative views about their own experiences as student bloggers, and reported that the blogging format did not motivate them to do any more than the bare minimum required by the teacher.

It is tempting to interpret these sets of positive and negative evaluations as indicators of the surface attitudes and true feelings of the participants in the study respectively; as indicating, in other words, that the students approved of the idea of BALL in principle, but did not actually like doing BALL activities in practice. However, the mode of analysis employed in this study provides no warrant for this explanation. From a phenomenological point of view, all statements about experience are equally ‘true’, and thus cannot be arranged into some kind of notional hierarchy of importance or authenticity. Nevertheless, the current study does allow us to advance some tentative explanations for why ESL students’ apparent enthusiasm for BALL does not translate into an upsurge of self-directed blogging activity.

For this group of learners, at least, enthusiasm for blogging in practice seems to have been inhibited by three interrelated factors. Firstly, the students found reading and writing blogs in a foreign language (in this case, English) an onerous task in itself, on account of their limited general English proficiency. In this sense, the students' experience of learning to write using a blog-assisted methodology seems to have been no different from the experience of learning to write in a more traditional ESL writing classroom context. Learning to write in a second language is hard work whatever the methodology, and the use of a blogging format does not seem to have made the learning process any easier, or sugared the pill to any appreciable degree.

Secondly, the participants in this study said they were inhibited by feelings of worry and embarrassment about uploading samples of their writing to the very public forum of an Internet blog. The main source of these anxieties seems not to have been the prospect of their work being viewed and evaluated by the teacher or even by members of the general public, but unsympathetic classmates who they perceived as likely to ridicule any errors they might make. The third inhibiting factor seems to have been the slow pace of the students' blogging activities. This factor is clearly related to the first; the students took a long time to complete each compulsory BALL task because their general English proficiency levels were quite low. As a consequence of this, the students were either unwilling or unable to devote any further time to non-compulsory blogging activities such as journaling or posting comments on peer work.

On the face of it, our finding that students do not find the blog format inherently any more motivating claimed by previous studies (e.g., Pinkman, 2005; Wu, 2008; Churchill, 2009; Ho & Usaha, 2009; Noytim, 2010; Trajtemberg & Yiakoumetti, 2011; Taki & Fardafshari, 2012) seems particularly disappointing. In retrospect, however, it is not entirely surprising that this should be so, given that the students in this research were aged between 18 and 19. This is the generation for whom Internet access and social media are as ubiquitous and unremarkable as the air that they breathe (Tan et al., 2010). The assumption that the learning motivation levels of these students may be enhanced by the use of Web 2.0 technologies may well therefore be a fundamentally misplaced one. Indeed, it probably tells us more about the perspectives of BALL researchers, most of whom will be from earlier generations who remember life before the Internet, and who may thus still regard Web 2.0 technologies as inherently wondrous and thus inherently motivating. If this is correct, however, it raises the intriguing possibility that BALL may be more motivating for older learners. This could be a fruitful line of inquiry for future BALL research to pursue.

The finding that the students were anxious about peer criticism constitutes a pedagogically much more tractable problem. Specifically, it may be ameliorated by the addition of an explicit learner training component to the BALL program, in which the students learn about the value and importance of receiving and responding to constructive criticism on their work. Following on from this, the course could also incorporate explicit instruction in how to write constructively critical feedback on other students' work. Thus, learner training input could become part and parcel of the core writing skills curriculum, rather than being seen as a peripheral add-on. The problem of slow task completion rates might also be dealt with in much the same way. It is clear from the interview data that the students in the current study did not really understand that while formal writing tasks were geared toward the development of grammatical accuracy and structural appropriacy, the goal of the journaling and peer commenting activities was to encourage and develop written fluency. They treated *all* blogging activities as accuracy-based activities, and consequently spent disproportionately large amounts of time completing what were supposed to be quick, short and informal blog posts. More explicit consciousness raising activities may have helped these students to understand more clearly the specific aims of the different kinds of writing task employed in the BALL format, and to adjust their own writing practices and expectations accordingly. Ensuring clarity about the form and function of the teacher feedback provided on different kinds of task is also crucial here. We would suggest that whereas it is entirely reasonable for students to expect formal essay writing tasks to be evaluated in terms of accuracy, it might be advisable for teacher responses to journal entries and other kinds of informal blog posting activities to focus on substantive content rather than lexicogrammatical form as a matter of general policy, and to inform students about this policy (and the rationale for it) at the very beginning of a BALL program. Paradoxically, given the widespread popular perception of blogging as form of personal diary writing, it may even be advisable to restrict BALL activities to formal assignments, and to revert to traditional paper notebooks as the medium for more personal and informal writing activities where the emphasis is on fluency and free expression rather than formal accuracy and reader evaluation. Alternatively, students could upload personal journal entries to an area of their blogs to which only they and their teacher have access, if the particular blogging software being used has this functionality.

It should also be noted that there are a number of key limitations to this study that prevent us from making an overall judgement of this methodology for the time being. First of all, it must be acknowledged that the research reported in

this thesis is based on just one of many possible teaching approaches. Specifically, students worked by themselves on most occasions (although they were encouraged to blog with each other for learning), and uploaded work on a frequent basis and after comparatively short periods of preparation and drafting. This raises the question of whether these conditions may or may not have contributed to the students' negative attitudes toward the blogging format in general. This is a question that can only be answered by repeating the current study using relatively more collaborative approach, in which students would typically be required to work collaboratively, and to upload work only after extensive periods of drafting and redrafting, and after extensive peer feedback sessions. It would be interesting to see whether this approach would lead to any enhanced learning experience compared to those reported in this thesis. Improving students' learning experience with BALL (particularly those of low confidence in themselves or of low proficiency levels in English) may be possible when teacher-researchers help students to overcome their shyness or mental obstacles of publishing entries publicly online. This can be done adopting a relatively slower blogging pattern or allowing more collaborative discussions. To be specific, teacher-researchers may consider getting students to publish a paragraph or short-length text expressing their initial thoughts first, as these are easier to manage and therefore mistakes in writing can be ironed out. After students develop more confidence or are used to exposing their thoughts in relatively public settings, teacher-researchers can move students to work on lengthier articles. Allowing students to have collaborative discussions or work as a pair/group may also be a way of alleviating students' anxieties about publishing their individual work and expressing their own ideas and feelings.

It must also be acknowledged that the results of the study may have been affected to some degree by the classroom setting, in which the students were working individually at PC terminals at all times during classes. Again, it would be useful for future studies to consider the impact of different forms of classroom organisation on the students' overall writing performance and/or perceptions of the BALL methodology. On a related note, it is also worth noting that the blogging approach may be more or less suitable for students who have different learning styles. This being the case, it would be interesting and useful for future studies to take students' learning styles into consideration when blog use is introduced into classrooms.

Thirdly, the students taught using the blogging approach in the study were often put off because of their limited English proficiency, which seemed to deter them from blogging or getting motivated to learn writing. It may be anticipated that a group of advanced EFL student bloggers may experience classroom blogging differently, as their higher proficiency levels in English may give them more control over blogging activities. This is also a question for further studies to investigate.

Finally, while our analysis does not lead us to conclude that the hype surrounding BALL is misplaced, it does suggest that more caution needs to be injected into the discussion. BALL is still a relatively new pedagogy and there remains much to learn about it. What is already clear from our research, however, is that BALL should not be seen as a 'magic bullet' solution for promoting learning motivation among second language writing students. The students in our study said that they liked the blog-based format, but in practice they did not find it any more (or any less) motivating than any other methodology. Our analysis also suggests that we should bear in mind the pace of technological change in the modern world, and how this affects the perspectives of different groups of observers: what seems exciting and thus intrinsically motivating to one generation may be humdrum and unremarkable to another. In the meantime, we would advise against jumping to any premature conclusions about whether second language writing teachers should run with the BALL, throw it to a team-mate, or drop it altogether.

## References

- Arani, A. J. (2005). *Teaching reading and writing English through a web-based communicative medium: Weblog*. ESP WORLD, 3(4). Retrieved 16 January, 2009, from [http://www.espworld.info/Articles\\_11/TeachingReadingandwritinginESPthroughaWeb-BasedCommunicativeMedium.htm](http://www.espworld.info/Articles_11/TeachingReadingandwritinginESPthroughaWeb-BasedCommunicativeMedium.htm)
- Arslan, R. Ş., & Şahin-Kızıl, A. (2010). *How can the use of blog software facilitate the writing process of English language learners?* Computer Assisted Language Learning, 23(3), 183–197.
- Campbell, A. P. (2003). *Weblogs for use with ESL classes*. The Internet TESL Journal, 9(2). Retrieved 15 July, 2009, from <http://iteslj.org/Techniques/Campbell-Weblogs.html>

- Cheng, M. C. (2006, March). Integrating the Internet into English composition class: Blogs, instant messaging, and e-mail. *Paper presented at 2006 International Conference and Workshop on TEFL and Applied Linguistics, Department of Applied English, Ming Chuan University, Taipei, Taiwan.*
- Chiao, T. L. (2006, July). Application of blogging in EFL writing class: A case study. *Paper presented at the Fifth International Symposium on English teaching, Taipei, Taiwan.*
- Churchill, D. (2009). *Educational applications of Web 2.0: Using blogs to support teaching and learning.* British Journal of Educational Technology, 40(1), 179–183.
- Dörnyei, Z. (2001). *Motivational strategies in the language classroom.* Cambridge: Cambridge University Press.
- Dörnyei, Z., & Ushioda, E. (2011). *Teaching and researching motivation* (2nd Ed.). Harlow: Pearson Longman.
- Fellner, T., & Apple, M. (2006). *Developing writing fluency and lexical complexity with blogs.* The JALT CALL Journal, 2(1), 15–26.
- Giorgi, A. P., & Giorgi, B. (2008). *Phenomenological psychology.* In Stainton-Rogers, W. & Willig, C. (Eds.), *The SAGE handbook of qualitative research in psychology* (pp. 165-178). London, UK: SAGE Publications.
- Ho, P. V. P., & Usaha, S. (2009). *Blog-based peer response for EFL writing: A case study in Viet Nam.* AsiaCall Online Journal, 4(1), 1–29.
- Hunter, L., Muilenburg, L., & Burnside, R. (2012). *Second language practice using blogs and pen pals with native speakers.* In T. Amiel & B. Wilson (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2012* (pp. 2137–2142). Chesapeake, VA: AACE.
- Hyland, K. (2003). *Second language writing.* New York, NY: Cambridge University Press.
- Iida, A. (2009). *Research in weblog pedagogy: Blogging and developing learner autonomy in a JFL context.* The Language Teacher, 33(2), 3–7.
- Kang, I., Bonk, C. J., & Kim, M. (2011). *A case study of blog-based learning in Korea: Technology becomes pedagogy.* Internet and Higher Education, 14(4), 227–235.
- Lin, M. H. (2012). *Blog assisted language learning in the EFL writing classroom. An experimental study. (Unpublished doctoral dissertation).* University of Birmingham, Birmingham, UK.
- Lin, M. H., Lin, C. Y., & Hsu, P. Y. (2011). *The unrealistic claims for the effects of classroom blogging on English as a second language, students' writing performance.* British Journal of Educational Technology, 42(6), E148–E151.
- McNamara, M. S. (2005). *Knowing and doing phenomenology: The implications of the critique of 'nursing phenomenology' for a phenomenological inquiry: A discussion paper.* International Journal of Nursing Studies, 42(6), 695–704.
- Miyazoe, T., & Anderson, T. (2010). *Learning outcomes and students' perceptions of online writing: Simultaneous implementation of a forum, blog, and wiki in an EFL blended learning setting.* System, 38(2), 185–199.
- Moustakas, C. (1994). *Phenomenological research methods.* Thousand Oaks, CA: Sage.
- Nguyen, P. T. (2012). *Peer feedback on second language writing through blogs: the case of a Vietnamese EFL classroom.* International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT), 2(1), 13–23. doi:10.4018/ijcallt.2012010102
- Noytim, U. (2010). *Weblogs enhancing EFL students' English language learning.* Procedia Social and Behavioral Sciences, 2(2), 1127–1132.
- Pinkman, K. (2005). *Using blogs in the foreign language classroom: Encouraging learner independence.* The JALT CALL Journal, 1(1), 12–24.
- Richardson, W. (2009). *Blogs, wikis, podcasts, and other powerful Web tools for classrooms* (2nd ed.). Thousand Oaks, CA: Corwin Press.
- Sayed, O. H. (2010). *Developing business management students' persuasive writing through blog-based peer-feedback.* English Language Teaching, 3(3), 54–66.
- Sun, Y. C. (2010). *Extensive writing in foreign-language classrooms: A blogging approach.* Innovations in Education and Teaching International, 47(3), 327–339.

- Sun, Y. C., & Chang, Y. J. (2012). Blogging to learn: Becoming EFL academic writers through collaborative dialogues. *Language Learning & Technology*, 16(1), 43–61.
- Taki, S., & Fardafshari, E. (2012). Weblog-based collaborative learning: Iranian EFL learners' writing skill and motivation. *International Journal of Linguistics*, 4(2), 412–429.
- Tan, K.E., Ng, M.L.Y., & Saw, K.G. (2010). Online activities and writing practices of urban Malaysian adolescents. *System*, 38(4), 548–559.
- Trajtemberg, C., & Yiakoumetti, A. (2011). Weblogs: A tool for EFL interaction, expression, and self-evaluation. *ELT Journal*, 65(4), 437–445.
- Yunus, M. M., Salehi, H., & Chenzi, C. (2012). Integrating social networking tools into ESL writing classroom: Strengths and weaknesses. *English Language Teaching*, 5(8), 42–48.
- Wang, H. C. (2009). Weblog mediated peer editing and some pedagogical recommendations: A case study. *The JALT CALL Journal*, 5(2), 29–44.
- Ward, J. M. (2004). Blog assisted language learning (BALL): Push button publishing for the pupils. *TEFL Web Journal*, 3(1). Retrieved October 25, 2008 from [http://www.teflWeb-j.org/v3n1/blog\\_ward.pdf](http://www.teflWeb-j.org/v3n1/blog_ward.pdf)
- Warlick, D. F. (2007). Classroom blogging: A teacher's guide to blogs, wikis, and other tools that are shaping a new information landscape. *Raleigh, NC: Landmark Project*.
- Warschauer, M. (2010). Invited commentary: new tools for teaching writing. *Language Learning & Technology*, 14(1), 3–8.
- Wu, W. S. (2008). Using blogs in an EFL writing class. In Priya, S. (Ed.), *Netlingo: The metamorphosis of language* (pp. 86–99). Hyderabad, India: The Icfai University Press.

## Prompts-based Scaffolding for Online Inquiry: Design Intentions and Classroom Realities

Meilan Zhang

Department of Teacher Education, University of Texas at El Paso, EDUC 801D, 500 W. University Ave., El Paso, TX, 79968 // mzhang2@utep.edu

(Submitted December 30, 2011; Revised May 26, 2012; Accepted August 16, 2012)

### ABSTRACT

With easy access to and exponential growth of online resources in the last decade, the Web has become the primary information source to middle school students for their school work. Yet, prior research has shown that students have difficulty evaluating, reading, and taking notes from online resources. This study analyzed the effectiveness of a digital notepad, which used prompts to scaffold middle school students in learning with online scientific resources. Data were collected from 8 sixth grade students who were engaged in a two-week online inquiry for a science project. Data analysis showed a gap between design intentions and classroom realities. Despite the prompts intended to promote students' critical evaluation of websites, their evaluation was still quick and dichotomous. In addition, students demonstrated different patterns in responding to the reading and note-taking prompts. However, prompts that aimed to promote deep thinking were generally answered with superficial responses. Implications for instruction and research involving use of online resources are discussed.

### Keywords

Online resources, Prompts, Evaluation, Note-taking, Middle school

### Introduction

The rapid expansion of Web access and abundance of online resources provide unprecedented opportunities for classroom teaching and learning (Wells & Lewis, 2006). According to recent national surveys (DeBell & Chapman, 2006; Johnson, Levine, Smith, & Stone, 2010), the majority of middle school students, who are among the largest and fastest-growing Internet user groups, rely on Web resources, rather than libraries for their school work. However, prior research has shown that middle school students have great difficulty using online resources effectively (Kuiper, Volman, & Terwel, 2009; Wallace, Kupperman, Krajcik, & Soloway, 2000).

One of the challenges that students face is to evaluate online information for quality and credibility. Unlike traditional books, publishing content online is easy and instant. Anyone with Internet access can publish virtually any content on the Web immediately. Therefore, incomplete or inaccurate information on the Web is not uncommon, which requires students to be critical consumers of online information. Yet prior research has found that middle school students either rarely spontaneously evaluate online information (Britt & Aglinskias, 2002; Walraven, Brand-Gruwel, & Boshuizen, 2009), or use naïve strategies to assess a site (Wallace, et al., 2000).

In addition, many middle school students read online resources in a superficial manner. They tend to look for quick, ready-made answers for their questions rather than developing meaningful content understanding (Wallace, et al., 2000). Many middle school students lack the reading strategies employed by skilled readers in online reading, such as reading with a clear purpose, monitoring reading comprehension, looking for evidence to support main ideas, or summarizing what they have learned after reading a website (Brozo & Simpson, 2002; Coiro & Dobler, 2007; Zhang & Duke, 2008).

Adding to the challenges is note-taking from online resources. Piolat, Olive, and Kellogg (2005) suggested that note-taking is more demanding than reading as students need to comprehend what they are reading, select important points, and record them. Typical problems in middle school students' note-taking from online resources include copying information verbatim with little thinking, manual recording on paper that is time-consuming and error-prone, and taking notes irrelevant to their research questions (Zhang & Quintana, 2012). Moreover, notes taken on paper are often disconnected to the original sources. When students forget to record the URL of a site in their notes, it is difficult to return to the original website.

Clearly, students need support to use the Web as an information source effectively. One of the scaffolding strategies often used in technology-enhanced learning environments is prompts. Prior research suggested that prompts can enhance metacognitive planning and reflection and improve content understanding (Davis, 2000, 2003; Quintana et al., 2004). Davis (2000, 2003) found that prompts embedded in a Web-based program called *Knowledge Integration Environment* helped students reflect on and synthesize science knowledge. Sandoval and Millwood (2005) integrated prompts in a software tool called *ExplanationConstructor* to guide high school biology students' construction of explanation concerning natural selection. Li and Lim (2008) found written prompts effective in scaffolding middle school students in online historical inquiry tasks. Ge and Land (2003) found question prompts had positive effect on undergraduate students' performances in solving an ill-structured problem. In the study of Walton and Archer (2004), an evaluative checklist prompted university students to consider "Who, why, what, when, where, and for whom" when using online academic resources. In addition, Quintana et al. (2004) recommended prompts and reminders as a software-based scaffolding strategy to support ongoing articulation and reflection during scientific investigations.

Based on the research literature that indicates the positive effects of prompts in technology-enhanced learning environments, it seems reasonable to assume that prompts can be useful for supporting student learning with online resources. We designed a software program called *Digital IdeaKeeper*, which embedded prompts to support students in online inquiry, a set of interconnected processes that include asking a research question, searching for information on the Web, evaluating and analyzing information, and synthesizing information to answer the research question (Kuiper, et al., 2009; Quintana, Zhang, & Krajcik, 2005). Typically, in online inquiry, students create final products such as essays, artifacts, brochures, and presentations to demonstrate their understanding gained from online resources.

Our prior research has found that IdeaKeeper helped students engage in deeper exploration of fewer websites, improved efficiency by focusing students' attention on learning goals and automatizing mechanical tasks, and enhanced metacognitive planning and monitoring (Zhang & Quintana, 2012). In another study, we found that IdeaKeeper-structured online reading was more deliberate, thorough, and purposeful than the one by students who did not use this tool (Zhang, in press). While IdeaKeeper attempts to support the full range of online inquiry activities including inquiry planning, information search, analysis, and synthesis, this study focused on the digital notepad within IdeaKeeper, which aimed to scaffold students in online evaluation, reading, and note-taking through prompts.

The purpose of this study was to understand how prompts as a specific software scaffolding strategy can support students in online inquiry in real-world classroom practice. Although researchers generally found prompts useful for promoting reflection and articulation in science inquiry, little is known about how middle school students actually respond to prompts designed to support evaluation, reading, and note-taking with online scientific resources.

## **Design of a digital notepad to support evaluation, reading and note-taking**

Quintana et al. (2004) suggested that making tacit disciplinary thinking explicit in software tools can help novice learners understand the steps they need to undertake in their work. Therefore, the design of the IdeaKeeper notepad highlights the implicit processes involved in learning with online information: evaluating, reading and summarizing (Schmar-Dobler, 2003; Zhang & Duke, 2008). Correspondingly, the digital notepad includes three tabs called "Skim," "Read," and "Summarize." In each of the tabs, prompts are embedded to guide students to evaluate, read, and take notes from online resources (See Figure 1). During the design processes, we worked with two experienced middle school science teachers to ensure the language used in the prompts is appropriate for middle school students. In addition, we conducted a pilot test to improve the usability of the software.

The notepad is embedded in a browser. When students read a website, they can use the digital notepad to take notes. Notes are saved on an online server and can be retrieved on any computer. The digital notepad automatically records the link where a note is taken, keeping the note closely connected to its original source, so students can conveniently revisit the original site.

In the "Skim" tab, an evaluative checklist is presented to help students assess the credibility and quality of online resources. The evaluative guide that we developed focuses on helping middle school students become mindful of the relevance of online resources with respect to their learning goals (Zhang & Duke, 2008), reading levels, differences

between facts and opinions (Kymes, 2005), date of publishing, author credibility and bias, and trustworthiness (Kuiper, et al., 2009). The evaluative prompts are as follows: (1) Is this website related to my question? (2) Is this website easy to read? (3) Is the information on this website up to date? (4) Is the information on this website fact or opinion? (5) Are the pictures or animations helpful or distracting? (6) Might the author be biased by presenting this information this way? (7) Does the author know what he or she is talking about? (8) How much do I trust this website? And (9) Is this website worth reading?

The figure displays three panels of a digital notepad interface, each with a different tab selected: 'Skim', 'Read', and 'Summarize'.

- Skim View (Left Panel):** The 'What I want to learn:' section has a bullet point: 'the reason why the sky is blue'. Below it are three tabs: 'Skim' (selected), 'Read', and 'Summarize'. The 'This website is:' section has radio buttons for 'related to my question', 'not related', 'easy to read', and 'hard to read'. The 'The information on this website is:' section has radio buttons for 'up to date', 'out of date', 'N/A', 'fact', 'opinion', and 'both'. The 'The pictures or animations are:' section has radio buttons for 'helpful', 'distractive', and 'N/A'. The 'The author:' section has radio buttons for 'might be biased by presenting this information this way', 'not biased', 'knows what he or she is talking about', and 'does not know'. The 'How much do I trust this website?' section has a scale from 1 to 5. The 'Is this website worth reading?' section has radio buttons for 'Yes' and 'No'.
- Read View (Middle Panel):** The 'What I want to learn:' section has a bullet point: 'the reason why the sky is blue'. Below it are three tabs: 'Skim', 'Read' (selected), and 'Summarize'. The 'What is the main idea of this webpage?' section has a large text area. The 'What evidence does the author give to support the main idea?' section has a large text area. The 'What information helps me answer my questions?' section has a large text area. The 'What information in this web page is difficult to understand?' section has a large text area. The 'What points does this author make that other authors do not?' section has a large text area. The 'What is other relevant information?' section has a large text area.
- Summarize View (Right Panel):** The 'What I want to learn:' section has a bullet point: 'what other countries dump raw sewage into the ocean?'. Below it are three tabs: 'Skim', 'Read', and 'Summarize' (selected). The 'What did I learn from this site?' section has a large text area. The 'You can refer to your notes below to write your summary.' section has a large text area. The 'What is the main idea of this webpage?' section has a large text area.

Figure 1. Notepad in “Skim,” “Read” and “Summarize” view

In the “Read” tab, a list of prompts is provided to foster critical and purposeful reading and note-taking. These prompts guide students to focus on major ideas rather than trivial details (Yang & Hung, 2008), connect reading to learning purposes (Zhang & Duke, 2008), monitor reading comprehension (Block & Pressley, 2002), and corroborate ideas from multiple sources (Strømsø, Bråten, & Samuelstuen, 2003). These prompts include: (1) What is the main idea of the webpage? (2) What evidence does the author give to support the main idea? (3) What information in the webpage helps me answer my questions? (4) What information in this webpage is difficult to understand? (5) What points does this author make that other authors do not? And (6) What is other relevant information?

In the “Summarize” tab, students are prompted to review “What did I learn from this website?” Research has shown that summarization can foster reading comprehension, helping learners integrate new information with what they already know and understand a text at a deeper level than simply reading it (Kymes, 2005; Palincsar & Brown, 1984).

The overarching question of this study was: How did middle school students use a prompts-based digital notepad to evaluate, read and take notes from online resources? In particular, three research questions were examined: (1) How did students use the evaluative prompts to evaluate online resources? (2) How did students respond to the reading and note-taking prompts? and (3) How did students demonstrate individual differences in using the notepad?

## Method

This study employed a multiple case study design to investigate the effectiveness of the digital notepad, which was used in an extended science inquiry project.



## Context

This study took place in a sixth grade classroom at a public middle school in a Midwestern city in the United States. The students were attending a 9-week writers' workshop where they were taught to improve writing skills. Two teachers taught the two classes in turn. One teacher was responsible for creative writing and another for scientific writing. Each topic lasted four and half weeks. This study occurred when the students were with their scientific writing teacher. For the scientific writing project, the teacher asked students to generate their own research question and sub-questions and then to search for information online. The students met for one hour in each weekday afternoon in the computer lab for two weeks. Afterwards, the students spent three days offline creating a brochure to demonstrate their understanding about their research question. Each student pair also presented their research to the class on the last day of the online inquiry project. Their teacher had 11 years of experience teaching language arts, science, and geography at the middle school level, and had integrated Web resources into her teaching for more than 8 years.

## Participants

Participants were 8 sixth grade students, including 6 Whites, 1 Asian, and 1 African American student. Prior to going online, the students generated research questions on water, which was the unit topic they were learning in their science class. They were grouped into pairs by their teacher according to the similarity of their questions. After being assigned into a pair, the students discussed their questions and reached consensus on the specific questions to study. The students used IdeaKeeper for the online inquiry task and the digital notepad in IdeaKeeper to take notes. They received training on how to use IdeaKeeper and spent one day practicing the tool before using it for their project. Table 1 describes the background of the student participants in this study. Their academic achievement information was provided by their teacher.

*Table 1.* Background of student participants

Dyad	Name	Gender	Achievement level	Research question
Dyad 1	Dane & Abby	Boy-Girl	Medium-High	How does acid rain affect the quality of our water?
Dyad 2	Clay & Gwen	Girl-Girl	High-High	Why are countries allowed to dump raw sewage in the ocean?
Dyad 3	Ferris & Edna	Boy-Girl	Medium-Medium	What should we do about 1,4 dioxane in Ann Arbor's groundwater?
Dyad 4	Dylan & Bona	Boy-Girl	Medium-Medium	How does a combination of groundwater and surface water affect the quality of Ann Arbor's drinking water?

## Data collection and analysis

Data sources for this study were the screen videos of the students' online activities and conversations captured by Camtasia, a screen recording program. The students spent 9 class sessions in the computer lab on their online inquiry task, 40-50 min per session. All of the sessions in the computer lab were captured. As a result, a total of 36 ( $4 \times 9 = 36$ ) screen videos—about 27 hours of videos—were collected from the students. Camtasia captured the screen with high resolutions, so the notes that the students took with the notepad were clearly recognizable.

Data analysis was an iterative process guided by the grounded theory approach (Glaser & Strauss, 1967). First, during the data collection period, the author made classroom observational notes for all the online inquiry sessions. Some initial patterns were observed and noted. For example, it appeared that the students generally had difficulty answering the bias prompt. After data collection, the author viewed and transcribed all of the videos collected from the students. The patterns of quick, dichotomous evaluation, vague responses to prompts, and individual differences gradually emerged during the long process of viewing and transcribing the videos. Analytic memos on emerging patterns were made for each pair.

The three patterns were further examined through detailed analysis on both the processes of using the notepad and the products (i.e., notes) that the students generated. In terms of the processes, the following aspects were analyzed: (1) Did the students compose their notes or copy and paste information? (2) In what order did the students respond to the prompts in the notepad? And (3) How long did the students spend on each prompt?

In addition, the student responses to the prompts in the *Skim*, *Read*, and *Summarize* tabs were compiled. The number of notes produced by each pair, the amount of words in their notes, and the content of their responses were analyzed. Moreover, content analysis was conducted on the students' conversions, which provided a window into their thinking related to the use of the digital notepad.

During the data analysis, both confirming and disconfirming evidence was examined. For example, although the students generally produced vague responses to the prompts, some pairs generated elaborated, specific responses to certain prompts. In this case, both general and particular patterns were reported. Also, data from different sources were examined to corroborate findings. For example, the pattern of quick and dichotomous evaluation of websites was corroborated by evidence from the students' conversions and their actual responses to the prompts. Their conversations showed that they tended to quickly categorize websites into "good" or "bad," which was consistent with the time they spent on the *Skim* tab and the nature of their responses to the evaluative prompts. The pattern of vague responses to the prompts was corroborated with their teacher's observations, who found vague responses common in the students' note-taking.

## Results

Data analysis revealed a gap between design intentions and classroom realities regarding the use of prompts to scaffold online inquiry. Despite the prompts intended to promote students' critical evaluation of websites, the students were still quick and dichotomous in evaluating websites—either "good" or "bad." They also tended to generate vague, nonmeaningful responses to the prompts in the reading and summary tabs. In addition, the students demonstrated different patterns in their use of the digital notepad. However, the prompts that were designed to promote deep thinking were generally answered with superficial responses. Next, the three findings were discussed, which answered the three research questions, respectively.

### Quick and dichotomous evaluation

Data analysis showed that the students tended to make dichotomous and emotional evaluation of a website—they either liked it or disliked it. When they liked a website, they used adjectives such as "good," "helpful," "cool," "awesome," "perfect," and "wonderful" to indicate their positive evaluation. When they disliked a website, they considered it "not good," "not helpful", or even "stupid," "boring," "sucks," "dumb," and "disgusting."

In addition, the students made their judgments about a site quickly. In general, it took fewer than 10 seconds for the students to decide whether they liked a site or not. The students would answer the evaluative prompts in the skimming tab if they decided a site was worth reading. Yet those prompts did not seem to engage the students in critical evaluation as intended. The students often went through the evaluative prompts quickly and spent fewer than 30 seconds answering the questions with little thinking. They relied on their first impression to evaluate a website, and often made self-repeating judgments in their conversations, such as, *"I think it looks up to date. It sounds like up to date. You know," "I think it is fact. It looks like fact, right?"* or *"I think they pretty much know what they are talking about. I think they ALL know what they are talking about."*

Each student pair evaluated websites individually. Together, the four pairs evaluated a total of 35 websites. Table 2 describes all the evaluative responses that the students generated for the 35 websites. One typical example was Dyad 4's evaluation for a website called "Hard water stops heart attacks" that they read on day 3. Their evaluation for this website was as follows: *"This website is related to my question. It is easy to read. The information is up to date. The information is fact. There are helpful pictures or animations in this site. The author is biased by presenting this information this way. The author knows what he or she is talking about. I trust this website. Yes, it is worth reading."*

Table 2. Summary of the students' evaluation of 35 websites

Evaluative prompts	Responses	Dyad 1	Dyad 2	Dyad 3	Dyad 4	Total
Is this website related to my question?	Yes	6	11	5	13	35
	No	0	0	0	0	0
Is this website easy to read?	Easy	6	11	5	7	29
	Hard	0	0	0	6	6
Is the information on this website up to date?	Up to date	1	8	5	5	19
	Not sure	4	1	0	7	12
	Out of date	1	2	0	1	4
Is the information on this website fact or opinion?	Fact	6	10	5	12	33
	Opinion	0	0	0	0	0
	Both	0	1	0	1	2
Are the pictures or animations helpful or distracting?	Not available	3	4	3	7	17
	Helpful	2	5	2	4	13
	Distractive	1	2	0	2	5
Might the author be biased by presenting this information this way?	Not biased	6	5	0	1	12
	Biased	0	0	5	10	15
	[No response]	0	6	0	2	8
Does the author know what he or she is talking about?	Yes	6	11	5	13	35
	No	0	0	0	0	0
How much do I trust this website?	1 (Not at all)	0	0	0	0	0
	2	0	0	0	2	2
	3	0	1	0	1	2
	4	1	2	1	9	13
	5 (Strongly trust)	5	7	4	1	17
Is this website worth reading?	Yes	6	11	5	13	35
	No	0	0	0	0	0

In particular, the conversations between the partners revealed that almost all the students had difficulty answering the prompt “Might the author be biased by presenting this information this way?” They did not know how to determine whether the author was biased or not. As a result, the students seemed to arbitrarily rate a site as either “biased” or “not biased.” For example, Dyad 1 rated all the websites they read as “not biased,” but Dyad 4 rated all sites as “biased” except for one website.

### Vague and nonmeaningful responses in note-taking

In general, the students produced vague and nonmeaningful responses to the prompts that intended to help them move beyond simply copying information from online resources, although some pairs were more thoughtful in note-taking than others. Overall, Dyads 3 and 4 generated less thoughtful responses to the prompts than Dyads 1 and 2. For example, the responses to the prompts in the reading and summary tabs in one of the notes that Dyad 4 produced were as follows. The questions in the square bracket refer to the prompts.

[What is the main idea of the webpage?] Health issues of hard water; [What evidence does the author give to support the main idea?] The author gives real examples of hard water; [What information in the webpage helps me answer my questions?] Hard water helps prevent heart attacks; [What information in this webpage is difficult to understand?] The information is not hard to understand; [What points does this author make that other authors do not?] The author includes information that has happened to support his or her statement; [What is other relevant information?] Most of the information is either on my topic or a real happening; and [What did I learn from this website?] Hard water prevents heart attacks.

In these responses, Dyad 4 focused on describing what the website was about, rather than specific science content. The only content-specific sentence in this note was: *Hard water helps prevent heart attacks*. Also, sometimes the students repeated their responses to different prompts. For example, in the note above the response to the summary prompt was identical to the response to prompt 3. Other pairs generated more or less similarly vague responses too. Next the patterns of how the students responded to each of the reading and summary prompts are described, which revealed a disparity between design intentions and classroom realities.

*Prompt 1: What is the main idea of the webpage?* This prompt intended to help students focus on the most important ideas rather than trivial details. However, the students seemed to interpret the prompt to be: *What is the webpage about?* The former required a statement, but the latter could be answered with one or two words. As a result, their responses to this prompt tended to be brief, for example, “the facts about acid rain,” “what is hard water made of,” and “the government of Panama.” These responses referred to the topic of a webpage, rather than the main idea, and did not convey specific, meaningful content.

*Prompt 2: What evidence does the author give to support the main idea?* This prompt aimed to promote scientific thinking when students read science-related websites because understanding how evidence supports a claim is essential in scientific reasoning. However, the student responses to prompt 2 suggested that simply asking students to identify evidence that supported the main idea was not enough to promote scientific thinking. Typically, the students gave a superficial, oversimplified answer to this prompt, such as, “They gave examples [to support the main idea],” “There are pictures or charts,” “There are big sub titles about the main topic,” or simply “They gave information.”

In addition, the student responses to prompt 2 often lacked a logical connection to their responses to prompt 1. For example, Dyad 1 considered one site’s main idea was to “tell us what acid rain is,” and evidence to support this main idea was “They give examples.” These responses showed that the students had difficulty understanding the concept of “evidence,” and how an idea was supported by evidence.

*Prompt 3: What information in the webpage helps me answer my questions?* This prompt was used heavily by Dyads 1 and 2, but Dyads 3 and 4 did not seem to treat this prompt differently from others, in terms of the amount of time they spent on this prompt and the number of words in their responses to this prompt. Dyad 1 perceived this prompt as most useful for their task and invested a great deal of time and effort in reading and taking notes for this prompt. Dyad 2 also copied a great amount of information to this prompt—964 out of 2,856 words in the 11 notes they took. Dyads 1 and 2 generated more specific and elaborated responses to prompt 3 than other prompts, while the responses of Dyads 3 and 4 to this prompt were still vague and short.

*Prompt 4: What information in this webpage is difficult to understand?* This prompt aimed to help students monitor their reading comprehension when they read a website. The student responses to this prompt fell into four categories: (1) Difficult concepts, such as Alkalies, pH level, and ammonia; (2) General comments about the content of the website, such as “This website is talking too scientifically;” (3) Interface features that made reading difficult, such as “The background is dark blue and so is the print;” and (4) The students claimed that the site was easy to understand, and nothing was difficult. Overall, the students’ monitoring of reading comprehension problems was limited to individual words. They knew when they did not understand a terminology, which was referred to as “big words” in their conversations. Beyond the word level, the students made only general comments about a site’s interface design (e.g., color or layout), and overall difficulty levels (e.g., easy or hard).

*Prompt 5: What points does this author make that other authors do not?* This prompt aimed to help students make connections among different sources they read. This prompt was challenging because it involved comparisons among different websites. To answer this prompt, the students had to recall what they had learned in previous websites and identify unique information from the current website being read. The students generated three types of responses to this prompt: (1) Specific new information provided by the current website (e.g., “Canada treats sewage before dumping it in the ocean”); (2) General differences or different interface designs (e.g., “This one gives lots of words that go with the topic and you can look up by clicking on it”); and (3) No differences (e.g., “nothing”). Overall, this prompt was less often answered by the students than other prompts.

*Prompt 6: What is other relevant information?* This prompt was complementary to prompt 3. It was a place where the students could save information that had not been covered by the other five prompts. Dyad 1 did not answer this prompt except for in one note. Dyads 2 and 3 who answered this prompt tended to copy and paste information, instead of composing a response. Dyad 4 tended to compose a quick answer to this prompt that was similar to other prompts. Examples of student responses to this prompt were as follows: “Look at the questions above” or “There is information about different types of [water] hardness.”

*Summary prompt: What did I learn from this website?* This prompt was designed to help students reflect on what they had learned from a website. Most students did use it to summarize and reflect on their learning in their own language. When the students answered the summary prompt, they typically re-read the content in the website and

referred to the notes they took in the reading tab. Some pairs wrote more elaborated and specific summaries than others, but the pattern of vague responses was still observed in some summaries. Overall, the students spent more time on the summary prompt and generated more elaborated answers for this prompt than most of the prompts in the reading tab.

#### Four profiles of notepad use

The students demonstrated a great deal of variety in using the digital notepad. The four dyads took 5, 11, 5, and 8 notes, respectively. They differed in terms of which prompts they focused on, copying (or composing) responses, and generating specific (or vague), elaborated (or short) responses in linear (or non-linear) order. Table 3 and Figures 2 and 3 describe the differences among the students. Four types of profiles of notepad use were observed.

Table 3. Patterns of composing or copying information for each prompt

	Dyad 1	Dyad 2	Dyad 3	Dyad 4
Prompt 1	Compose	Mainly Compose	Compose	Compose
Prompt 2	Compose	Mainly copy & paste	Compose	Compose
Prompt 3	Compose	Mainly copy & paste	Compose	Compose
Prompt 4	Compose	Compose	Compose	Compose
Prompt 5	Compose	Compose	Compose	Compose
Prompt 6	Compose	Mainly copy & paste	Mainly copy & paste	Compose
Summary	Compose	Compose	Compose	Mainly copy & paste

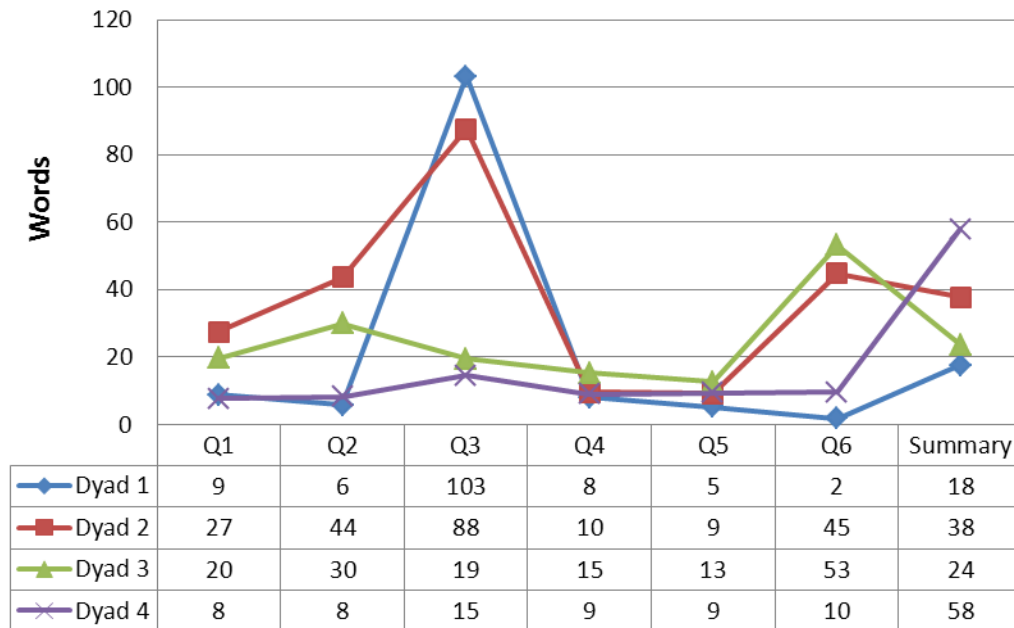


Figure 2. Number of words for each prompt per note

*Efficient functionalists.* Dyad 1 was considered efficient functionalists who viewed only prompt 3 useful for collecting information for their research question, while treating all other prompts as “assignments,” which they had to answer dutifully, but did not see the benefits of doing so. They were much more enthusiastic about prompt 3 than other prompts, and invested major efforts in answering this prompt, while trying to get other prompts done with minimal efforts. Their reading of websites was heavily centered on prompt 3—they did not read a webpage until they reached this prompt. When answering prompt 3, they often repeated the processes of “reading—note-taking—reading—note-taking” many times. While for other prompts, they typically scanned a webpage and wrote a quick answer. Dyad 1 composed a total of 752 words in all of the five notes they took, among which 516 words (69%) were composed for prompt 3. Overall, Dyad 1 spent 34 min (69%) taking notes for prompt 3, 7 min summarizing

what they had learned, and only 8 min on the other five prompts in the reading tab. Their responses to prompt 3 were much more specific, concrete, and meaningful than the “assignment” prompts.

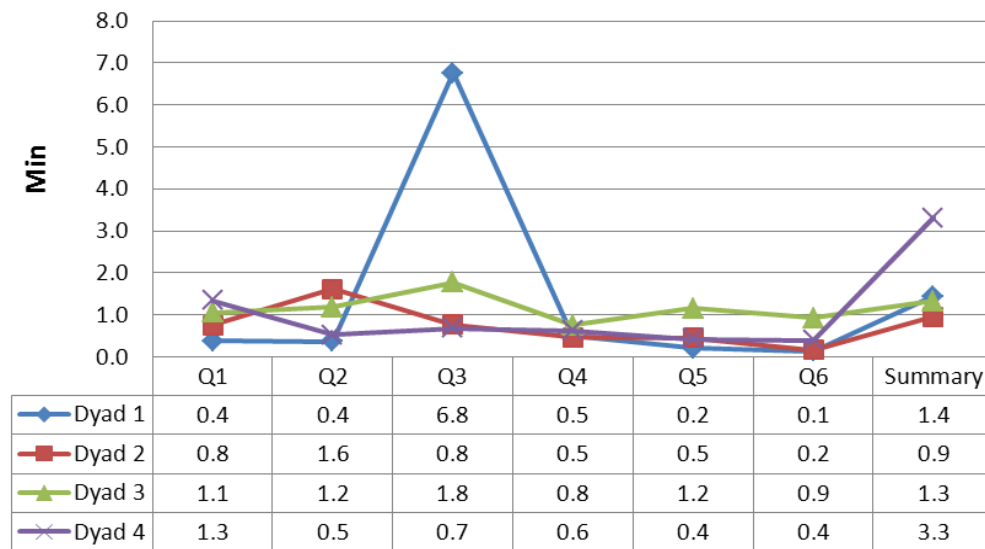


Figure 3. Time (in minute) on each prompt per note

*Information collectors.* Dyad 2 copied and pasted a great amount of information to the digital notepad. They were the only pair who copied and pasted information for most of the prompts, while the other three pairs mainly composed their notes. Overall, they copied and composed a total amount of 2,856 words, more than three times of words taken by the other pairs. The number of notes they took (11) was greater than the other three pairs too (5, 5, and 8). They copied the largest amount of words to prompt 3 (See Figure 2), although they did not spend more time on this prompt (See Figure 3). While Dyad 2 copied information in the reading tab, they composed summaries in their own language and referred to the notes they took in the reading tab. Thus the summary prompt provided an opportunity for Dyad 2 to enhance their understanding of the content they read in the websites.

*Hesitant vague repeaters.* Dyad 3 often repeated similar, vague responses to different prompts. For example, in one of the notes they took, their response to the first prompt was that “The main idea of this webpage is what hazards 1,4 dioxane can bring”, which was repeated in their summary: “What I learned from this website is ... what hazards 1,4 dioxane can bring.” In addition, Dyad 3 tended to focus on details and spent considerable time formatting their notes. For example, on day 2 they spent about two minutes simply changing the format of the term “C4 H8 O2” in their note. They were hesitant note-takers in the sense that they often typed some words, deleted them, and then typed some other words.

*Undifferentiated linear minimalists.* Unlike the other three pairs who all showed somewhat non-linear patterns in answering the prompts, for example, starting from prompt 3 or prompt 6 and returning to other prompts, Dyad 4 strictly followed the order of prompts in their note-taking. They tended to go through prompt 1 to prompt 6 and then the summary prompt linearly. Dyad 4 spent almost equal time answering different prompts in the reading tab and showed a tendency of getting work done with only minimal efforts. When answering the prompts, they tended to compose a quick answer without referring to the content in the webpages. Unlike other pairs who often revised their notes and added new content, Dyad 4’s note-taking was one-shot deal—they rarely returned to the prompts to revise the notes. Dyad 4 spent more time on the summary prompt, yet they tended to copy and paste information rather than composing their own responses.

## Conclusions

This study revealed the complexity in scaffolding students’ use of online resources with technological tools. Although the design of the digital notepad was grounded in relevant literature and attempted to address the problems

students encountered in evaluating, reading, and taking notes from online resources, the results of its use did not reflect the design intentions. Prior research has mainly reported positive results associated with using prompts in technology-enhanced learning environments (Davis, 2000, 2003; Ge & Land, 2003; Li & Lim, 2008; Walton & Archer, 2004). However, this study found that students can respond in a superficial manner to the prompts designed to promote critical evaluation and thoughtful reading and note-taking.

In particular, Papadopoulos, Demetriadis, Stamelos, and Tsoukalas (2011) found that undergraduate students who produced written answers to prompts achieved better performances in a Web-based learning environment than students who only thought of possible answers to the prompts and students who did not receive any prompts. Yet this study found that middle school students did not necessarily generate meaningful responses to prompts even when they had to produce written answers. One possible explanation is that students in this study were younger and needed substantial support and practice. Two important lessons can be drawn from this study.

First, deep learning on the Web is hard. It is one thing to know how to use the functions of a technological tool, yet another to use the tool in a pedagogically meaningful manner. Although all of the students in this study learned to use IdeaKeeper quickly, they did not see enough examples of good and poor responses to the prompts. As a result, they generated superficial responses to the prompts that aimed to promote deep thinking. The difficulty of facilitating deep scientific thinking is consistent with prior research (Oliver & Hannafin, 2000; Sandoval, 2003; Sandoval & Millwood, 2005). Sandberg and Barnard (1997) found that college students, when working with a computer-supported tutoring program, tended to get work done with minimal efforts and did not spontaneously engage in cognitive processes that led to deep learning. Sandberg and Barnard suggested that coordinated, sustained efforts by learners, teachers, and instructional designers are needed to promote deep learning. However, such a holistic approach is often lacking in studies on technology-enhanced learning environments. This study reinforced the importance of the holistic approach and suggested that teacher modeling, student practice, and timely feedback are important when introducing new technological tools in classrooms.

Second, individual differences add to the complexity of technological scaffolding for online inquiry. Profile analysis of students' online behaviors showed that students demonstrated great variety from each other, but often repeated the same individual patterns when working with different websites. The stability within individual students and the variety among different students have important implications for guiding the design of online learning environments.

In addition, the finding that most students had difficulty identifying "bias" is worrisome because of poor quality and discredited information abundant on the Web. Therefore, emphasis should be placed on developing students' awareness and skills in critically evaluating websites. Teachers should model how to differentiate facts from opinions and how to identify biases in websites, and provide adequate opportunities for students to practice.

Although the results were unexpected, this study still yielded important insights for designing technology-enhanced learning environments to scaffold students' online inquiry. First, this study added to understanding of how middle school students learn with online resources through the findings of quick and dichotomous evaluation, and vague and nonmeaningful responses in note-taking. Second, the detailed analyses of student responses to each prompt and descriptions of student profiles have implications for other researchers and designers who are interested in using prompts to scaffold online inquiry. It is often difficult for instructional designers to anticipate how learners respond to a scaffolded learning environment. This study suggested that students may interpret design features in ways different from what the designers expect. Future research should explore whether appropriate modeling and practice can improve the effectiveness of prompts-based scaffolding implemented in this study.

## **Acknowledgements**

This material is based on work supported by the National Science Foundation under Grant No. DUE-0226241. Any opinions and findings expressed in this material are those of the author and do not necessarily reflect those of the National Science Foundation. The author appreciates the invaluable contributions of the Digital IdeaKeeper research team, including Dr. Chris Quintana, Minyoung Song, Srikanan Reddy, and Allen Xiang.

## References

- Block, C. C., & Pressley, M. (2002). *Comprehension instruction: Research-based best practices*. New York, NY: Guilford Press.
- Britt, M. A., & Aglinskias, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction*, 20(4), 485-522.
- Brozo, W. G., & Simpson, M. L. (2002). *Readers, teachers, learners: Expanding literacy across the content areas* (4th ed.). Upper Saddle River, NJ: Merrill Prentice Hall.
- Coiro, J., & Dobler, E. (2007). Exploring the online reading comprehension strategies used by sixth-grade skilled readers to search for and locate information on the Internet. *Reading Research Quarterly*, 42(2), 214-257.
- Davis, E. A. (2000). Scaffolding students' knowledge integration: Prompts for reflection in KIE. *International Journal of Science Education*, 22(8), 819-837.
- Davis, E. A. (2003). Prompting middle school science students for productive reflection: Generic and directed prompts. *The Journal of the Learning Sciences*, 12(1), 91-142.
- DeBell, M., & Chapman, C. (2006). Computer and Internet use by students in 2003 (NCES 2006-065). Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Ge, X., & Land, S. M. (2003). Scaffolding students' problem-solving processes in an ill-structured task using question prompts and peer interactions. *Educational Technology Research and Development*, 51(1), 21-38.
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. New York, NY: Aldine.
- Johnson, L., Levine, A., Smith, R., & Stone, S. (2010). *The 2010 horizon report*. Austin, TX: The New Media Consortium.
- Kuiper, E., Volman, M., & Terwel, J. (2009). Developing Web literacy in collaborative inquiry activities. *Computers & Education*, 52(3), 668-680.
- Kymes, A. (2005). Teaching online comprehension strategies using think-alouds. *Journal of Adolescent & Adult Literacy*, 48(6), 492-500.
- Li, D. D., & Lim, C. P. (2008). Scaffolding online historical inquiry tasks: A case study of two secondary school classrooms. *Computers & Education*, 50(4), 1394-1410.
- Oliver, K., & Hannafin, M. J. (2000). Student management of Web-based hypermedia resources during open-ended problem solving. *Journal of Educational Research*, 94(2), 75-92.
- Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1(2), 117-175.
- Papadopoulos, P. M., Demetriadis, S. N., Stamelos, I. G., & Tsoukalas, I. A. (2011). The value of writing-to-learn when using question prompts to support web-based learning in ill-structured domains. *Educational Technology Research and Development*, 59(1), 71-90.
- Piolat, A., Olive, T., & Kellogg, R. T. (2005). Cognitive effort during note taking. *Applied Cognitive Psychology*, 19(3), 291-312.
- Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., . . . Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *Journal of the Learning Sciences*, 13(3), 337-386.
- Quintana, C., Zhang, M., & Krajcik, J. (2005). A framework for supporting metacognitive aspects of online inquiry through software-based scaffolding. *Educational Psychologist*, 40(4), 235-244.
- Sandberg, J., & Barnard, Y. (1997). Deep learning is difficult. *Instructional Science*, 25(1), 15-36.
- Sandoval, W. A. (2003). Conceptual and epistemic aspects of students' scientific explanations. *Journal of the Learning Sciences*, 12(1), 5-51.
- Sandoval, W. A., & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction*, 23(1), 23-55.
- Schmar-Dobler, E. (2003). Reading on the Internet: The link between literacy and technology. *Journal of Adolescent & Adult Literacy*, 47(1), 80-85.



- Strømsø, H. I., Bråten, I., & Samuelstuen, M. S. (2003). Students' strategic use of multiple sources during expository text reading: A longitudinal think-aloud study. *Cognition and Instruction*, 21(2), 113-147.
- Wallace, R. M., Kupperman, J., Krajcik, J., & Soloway, E. (2000). Science on the Web: Students online in a sixth-grade classroom. *The Journal of the Learning Sciences*, 9(1), 75-104.
- Walraven, A., Brand-Gruwel, S., & Boshuizen, H. P. A. (2009). How students evaluate information and sources when searching the World Wide Web for information. *Computers & Education*, 52(1), 234-246.
- Walton, M., & Archer, A. (2004). The Web and information literacy: Scaffolding the use of Web sources in a project-based curriculum. *British Journal of Educational Technology*, 35(2), 173-186.
- Wells, J., & Lewis, L. (2006). *Internet access in U.S. public schools and classrooms: 1994-2005*. U.S. Department of Education. Washington, DC: National Center for Education Statistics.
- Yang, Y.-F., & Hung, Y.-C. (2008). The impact of externalizing readers' mental representation on the comprehension of online texts. *Educational Technology & Society*, 11(4), 246-264.
- Zhang, M. (2013). Supporting middle school students' online reading of scientific resources: Moving beyond cursory, fragmented, and opportunistic reading. *Journal of Computer Assisted Learning*, 29(2), 138-152.
- Zhang, M., & Quintana, C. (2012). Scaffolding strategies for supporting middle school students' online inquiry processes. *Computers & Education*, 58(1), 181-196.
- Zhang, S., & Duke, N. K. (2008). Strategies for internet reading with different reading purposes: A descriptive study of twelve good internet readers. *Journal of Literacy Research*, 40(1), 128-162.

## The Design of Social Agents That Introduce Self-reflection in a Simulation Environment

Pei-Lan Lei<sup>1</sup>, Sunny S. J. Lin<sup>2</sup>, Dai-Yi Wang<sup>3\*</sup> and Chuen-Tsai Sun<sup>1</sup>

<sup>1</sup>Department of Computer Science, National Chiao Tung University, Taiwan // <sup>2</sup>Institute of Education, National Chiao Tung University, Taiwan // <sup>3</sup>Department of Computer Science and Communication Engineering, Providence University, Taiwan // pennylei0419@gmail.com // sunnylin@faculty.nctu.edu.tw // ophelia.wang@gmail.com // ctsun@cis.nctu.edu.tw

\*Corresponding author

(Submitted June 10, 2012; Revised August 31, 2012; Accepted September 12, 2012)

### ABSTRACT

This study focuses on the design of several social agents that are intended to collect the self-reflections of learners while learners are immersed in simulation activities for knowledge building. The design of the agents follows 5W principles and seeks to encourage learners to expend mental effort upon multi-faceted learning and self-reflection. Using semantic networks, we developed dialogue lines for reflection-prompting agents. We analyzed the participants' answers using natural language processing technology to classify the sentences into positive and negative rankings. A preliminary field study with 117 high school students was conducted over three weeks to test the effects of agent-prompted self-reflection. The results demonstrated that 96% and 62% of participants separately completed the first and the second simulation activities (including the agent-prompted self-reflections respectively). Those who did not finish the activities were generally limited by time restrictions rather than a lack of motivation, as the participants typically considered the interactions with the agents to be interesting. The self-reflections elicited through the agent interviews were consistent with the reflections obtained from paper-pencil questionnaires and appeared to be stable over time. Future study, including investigations using a randomized experimental design with a control group, is needed to fully assess the effects of agent-prompted self-reflection.

### Keywords

Simulation, Intelligent agent, Self-reflection, Multiple intelligence

### Introduction

In Taiwan, there are high-stakes entrance examinations to high school (after 9<sup>th</sup> grade) and college (after 12<sup>th</sup> grade). However, the educational system has undergone a gradual transformation over the past decade; in particular, the system is evolving toward a process in which student placement is less dependent on mandatory university screening based solely on college entrance examination scores but instead incorporates greater flexibility for students to autonomously select educational options based on their personal values and motivations. Self-awareness has therefore become an increasingly critical aspect of student decisions regarding majors and careers. Over time, the curriculum structure and activities offered by Taiwanese high schools have incorporated many elements that seek to elevate students' self-understanding. However, these students often experience issues related to confusion during the process of self-exploration, immaturity, and difficulty with commitment, and as a result, a high proportion of freshmen do not enjoy their educational choices and display less adaptive symptoms in their college lives (e.g., procrastination, absence, or failing grades in their major courses, Ministry of Education, 2012). This observation leads us to regard self-awareness as being critical for everyone (students especially) as part of both the hygiene aspect of active learning (e.g., the regulatory process or self-directed learning, Gibbons, 2002) and the prevention of inadaptable learning (e.g., avoidance or loss of motivation).

Self-reflection is very likely to enhance the accuracy of situation-, task- and self-awareness that are in turn critical components in self-regulation learning (de Bruin, Rikers, & Schmidt, 2005; Greene & Azevedo, 2007). Self-reflection provides the only window into an individual's inner state. Unfortunately, this window is often foggy, especially when individuals report on experiences that span over longer timeframes (Schwarz & Sudman, 1994). Memory is imperfect and susceptible to bias. Some biases can be attenuated through the use of better interviewing techniques (Schwarz & Sudman, 1994). Real-time self-reflection may also avoid the weakness of memory biases (people tend to remember experiences with intense emotions and global experiences rather than specific details (Redelmeier & Kahneman, 1996; Wang, 2001)). Moreover, self-reflection in simulated problem-solving situations

can avoid situational biases (retrospective self-reflection is often conducted in a situation very different from the situation that we are interested (Schwarz, 1999, 2007)) and the use of global heuristics (Bank, Dishion, Skinner, & Patterson, 1990). In sum, if activities of self-reflection could be blended with the simulation tasks or situation, the resistance of this repetitive mindful works could be reduced. Therefore, we planned to introduce social agents to prompt multi-faceted self-reflection activities (based upon multiple intelligences (MI), Gardner, 1983) through students' active explorations in a simulation environment.

Previous research has proposed that simulation-based e-learning is an effective knowledge-building approach (Eskrootchi & Oskrochi, 2010; Liu, 2010; Padiotis & Mikropoulos, 2010). Simulations have been demonstrated to be an effective teaching and learning approach in the fields of science (Chen & Howard, 2010; Meij & de Jong, 2006), medicine (John, 2007; Rosen, 2008), engineering (Chen, Hong, Sung, & Chang, 2011; Hsieh & Sun, 2007), social science (Cuenca López & Martín Cáceres, 2010; Liu, 2010), and culture (Peterson, 2010; Ranalli, 2008). Researchers (Chen et al., 2011; Chou, Chan & Lin, 2003; Tan, Tse, & Chung, 2010) have suggested that the benefits of simulations include the dynamic provision of multiple representations, visualization, immersion scenarios, hands-on experiments/manipulations, rich feedback, multiple interactions, and, perhaps most importantly, the engagement of an individual student as an active agent in knowledge building. In simulations, the learner is always in control of the learning process; thus, the regulation of simulation-based e-learning is more critical than the regulation of other instructional approaches that are subject to greater teacher control. Within simulations, problem-solving scenarios, active operations/manipulations, and the provision of abundant feedback/prompts during the exploration process provide ample opportunities to understand the self-competence, cognition, preference, and affection (Chen & Howard, 2010; Ifenthaler, 2012) that shape the core of self-awareness.

The extant studies that have demonstrated how simulation environments and agents can prompt self-reflection are limited; therefore, we developed several principles to guide our design of tasks and agents and then implemented the system. To assess the function of the agent-prompted reflections, a field test was conducted as a preliminary analysis of the degree to which the design of agents embedded in a simulation environment successfully inspired student reflections. Because self-reflection data are typically collected through lengthy self-report questionnaires, we used data collected in this manner to judge the effectiveness of the agent interview. Participants engaged in two simulation activities, three agent-prompted MI self-reflections (pre-simulation, during the 1<sup>st</sup> simulation activity, and during the 2<sup>nd</sup> simulation activity), and two paper-pencil MI questionnaires (pre- and post-simulation). Our first aim was to observe whether students successfully finished the simulation tasks and the agent-prompted reflections. Students then compared the MI reflections collected through the agent interviews (situation-specific index) to the paper-pencil questionnaires (situation-general index) to explain which reflection method they preferred to report on their experiences. Finally, the relationships between the agent-prompted reflections and the questionnaires and between the pre- and post-simulation reflections were examined to determine the validity and reliability of the agent-prompted reflections. We sought to answer the following questions.

- How many students successfully finished all of the simulation tasks (including the agent-prompted reflections) within two 120-minute sessions?
- Which reflection method do learners perceive more favorably, agent prompting or a paper-pencil questionnaire?
- For participants that successfully finished the simulation activities and self-reflections, are the reflections related to the five MI aspects that were collected through the agent interviews consistent with those collected through the paper-pencil questionnaires? Do MI reflections, collected through either paper-pencil questionnaires or agent interviews, change over time or remain stable?

## **Literature review**

### **Simulations**

Computer-mediated simulation involves the construction of a computer-mediated environment in which human users can interact as in a real environment, allowing researchers/educators to observe and analyze social relations and collective patterns (Lin & Sun, 2003). Internet-mediated simulation is one kind of computer-mediated simulation. As the Internet supports communication and global access to multimedia information resources, Internet-mediated simulation enables real-time interaction among multiple participants (Martin, 2006).

Simulations are widely used in education because they fulfill the following important functions. (a) Observation and

visualization with multiple representations: When the pace of a social or scientific phenomenon is sufficiently fast, slow, large or small that the phenomenon is barely observable in natural situation, simulations can provide vivid observational perspectives (Chen et al., 2011; Meij & de Jong, 2006); in addition, simulations are capable of incorporating multiple representations of information (e.g., verbal and visual-spatial) to enhance learning outcomes for students with various needs. (b) Hands-on activities: Simulations provide learners with the ability to take the initiative in processes of exploration and discovery. Learners can alter the parameters or factors and thus have the flexibility to test the hypotheses that they generate, design scientific experiments, and practice interpreting the data collected from these simulative experiments (Chen et al., 2011; Eskrootchi & Oskrochi, 2010; Padiotis & Mikropoulos, 2010). (c) Feedback: During exploration, simulations can offer immediate, rich and authentic feedback that is beneficial for regulation. (d) Interaction: simulations allow learners to role-play and support multiple interactions with either real humans (such as peers and teachers) or agents. The partial structuring of the students' interactions with simulations can increase learning effectiveness (Eskrootchi & Oskrochi, 2010).

### **Intelligent agents**

Agents are important tools in pedagogical simulations. In multimedia learning systems, animated pedagogical agents perform special, critical instructional functions (Chou et al., 2003; Atkinson, Mayer, & Merrill, 2005; Hsieh & Sun, 2007). Agents can guide the process of knowledge construction either by providing explanatory feedback and highly contextualized problem-solving advice (Moreno & Mayer, 2005) or by eliciting students' thoughts and reflections (Huang, Yeh, Li, & Chang, 2010; Wu & Looi, 2012). They help learners attend to, understand, and deeply process instructional messages (Atkinson et al., 2005; Wu & Looi, 2012). Agents also increase the level of multiplicity in social learning environments and broaden community diversity (Chou et al., 2003). They create virtual relationships with the learners, speak with or respond to learners, and encourage student communication (Atkinson et al., 2005). Pedagogical agents may play authoritative or non-authoritative roles, including the following (Chou et al., 2003; Hsieh & Sun, 2007; Huang et al., 2010; Wu & Looi, 2012): (a) knowledge suppliers – agents are responsible for presenting their stores of knowledge (e.g., tutors or instructors); (b) learning companions – agents act as collaborators or competitors who make learning activities less repetitive and more diversified; (c) tutees – learners act as tutors who teach the agent tutees in a learning-by-teaching environment; (d) consultants – agents provide advanced suggestions; and (e) guides – agents guide learners or help them adjust to environments according to the learning progress.

Intelligent agents can gather information from the environment and interact with it; to achieve this goal, these agents require artificial intelligence (AI) technology to reason and make inferences that permit them to act autonomously. A semantic network is one type of knowledge representation for explaining and simulating human intelligent behavior. By analyzing topics in terms of their concepts and relationships, one can quickly pinpoint how one concept might depend on another, what must be previously known about the topic, and a logical sequence in which topics should be taught. Carbonell (1970) and Mauldin (1994) used a semantic network technique to simulate a Socratic tutor; this type of modeling provided adaptive feedback. Furthermore, natural language processing is used to process human language to facilitate communication between agents and users (Chou et al., 2003). The intention is to transform the human-computer interactions as authentic as human-to-human conversation.

### **Multiple intelligences**

Simulations can provide interdisciplinary components that induce students to integrate various materials. Therefore, simulations can provide learners with ample opportunities to practice the competences described in Gardner's Multiple Intelligences scheme. Gardner (1983) suggests that human intelligence incorporates the following features: (a) the ability to solve real-life problems; (b) the ability to generate new problems to solve; and (c) the ability to make something or offer a service that is valued within one's culture. From the analysis of the competence losses in patients with brain injuries, Gardner discriminates between eight intelligences: linguistic, logical-mathematical, spatial, bodily-kinesthetic, musical, interpersonal, intrapersonal and naturalist.

The MI strengths self-checklist allows learners to conduct self-reflection and identify which types of intelligence are their fortes (Stanford, 2003). Gardner (1993) believes that the best way to measure learner intelligence is through a realistic appraisal of an individual's performance in many types of associated tasks, activities and experiences.

Notation systems can be used to record observations of students that are operating a machine or dealing with disputes, either under real-world conditions or in virtual simulations (Gardner, 1998; Stanford, 2003).

## The simulation system

Our goal is for students to have constant opportunities to reflect on both self- competence/capability and preference/affection in enjoyable and natural conversations with social agents that are presented as human-like figures in simulation tasks. In this study, the simulation system allows learners to naturally acquire knowledge from simulated tasks and encourages learners to engage in a doing process to increase the probability of positive transfer and the contribution to meaningful learning. All of the scenarios require manager agents, task controller agents, partner agents, and reflection-prompting agents. Manager, task controller, and partner agents play various roles in the tasks/activities and interact with learners to support task completion. For example, agents may provide instructions/hints/suggestions for solving the tasks, present task results to learners, analyze learner data, control the script procedures, and encourage learners. The reflection-prompting agents play fictitious interrogator roles to prompt student reflections via elaborative interrogation techniques. They then send the self-reflection data to the manager agents for further analysis and notify the task controller agents that the student has finished the self-reflection. The architecture of our system is presented in Figure 1.

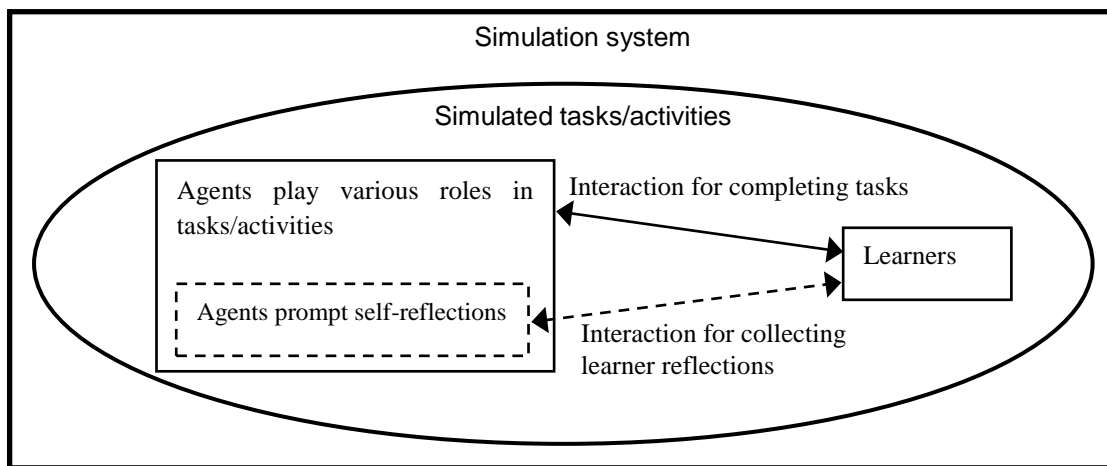


Figure 1. The architecture of our simulation system (the major design of this study is marked with the dotted line)

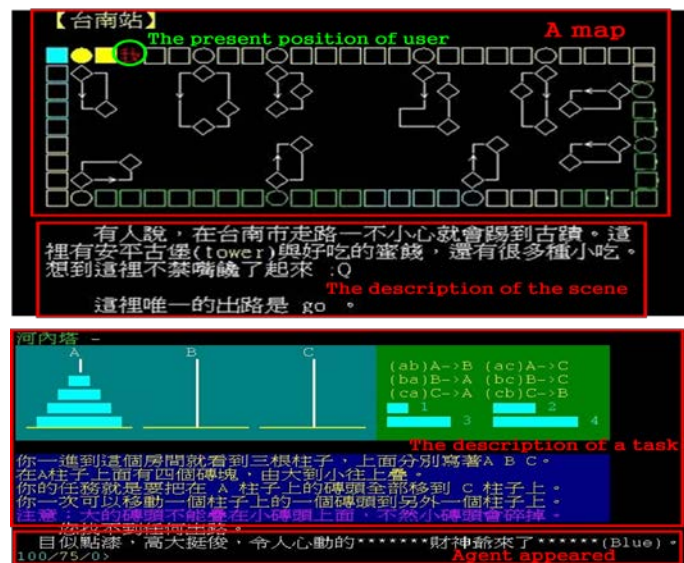


Figure 2. Sample user interface screen.

The platform of this study is a simulation system. The system has the following general characteristics. (a) It is Internet-based, containing multiple users playing roles in a virtual environment. (b) Users are introduced to the virtual room with a detailed description and illustration of the foreground and background environments. The user interface is presented in Figure 2. (c) Important attributes that determine user status include gender, age, energy and the amount of available money. (d) An online instruction manual is offered to help users with this system. (e) Users can type in text to communicate with other users or agents. (f) The simulation system can support collaborative learning; this study requires participants to collaboratively accomplish several tasks but individually accomplish self-evaluations. (g) Logs of learner behavior are saved in the system, allowing for further analysis by researchers.

### **The task design**

Before it introduces MI self-reflection activities, the simulation system provides interesting and diverse scenarios and agents that support hands-on tasks that allow individuals to exercise five types of MI via role-playing activities. In the field test described in this study, participants had two 120-minute sessions to explore the simulation environment and undertake the problem-solving tasks. Certain of these tasks require linguistic intelligence for reading and writing, because the tasks, scenes and characters are described using Chinese words and phrases. Learners have to read written instructions carefully to complete task requirements. Moreover, learners only communicate with each other or agents through writing and are sometimes required to write a short essay or journal. Several of the tasks require logical-mathematical intelligence, such as the “Tower of Hanoi” problem and the bucket-of-water task (learners have to use 3- and 5-liter buckets to obtain 1 liter of water). Certain tasks require spatial intelligence, such as reading maps and taking bearings to move freely among the scenes in the simulation environment. Certain tasks require interpersonal intelligence, such as asking learners to invite three or four people to work cooperatively to discuss strategies, overcome barriers, and collect treasures. Other tasks require intrapersonal intelligence, providing agents to promote the divergent thinking of learners. For example, these agents may ask learners, “Is doing so all right?”, “Will that be ok for you?”, and “How about doing it another way?”

The environment vividly simulates real-life situations. For example, in train tour scenarios (see Table 1), individuals can enter a train station, get on or off the train, exit the station, and visit local tourist attractions. Learners are asked to solve problems in a variety of situations, such as hunting for treasure in a train station or asking for help (collecting money) from a human partner to buy a train ticket. During the course of these tasks, the interrogator agents are responsible for prompting MI self-reflections. In certain tasks, learners can earn extra points or rewards or jump to the next stage faster as incentives for answering the self-reflection questions.

### **The design principles of the reflection-prompting agents**

In our system, self-reflection is elicited through casual and enjoyable conversations with bizarre, interesting agents in attractive role-playing scenarios. The reflection-prompting agents are committed to preserving enjoyment and social presence in interactions; therefore, they must be immersed within story scenarios and immediately communicate with learners. The agent-learner dialogues that are collected are further transformed by the system. The agent design follows 5W principles, as explained below.

- The “Who” principle: To maintain consistency across learning-reflection situations, the roles that the reflection-prompting agents play in the simulation environment are natural to the scenario. For example, an agent may announce her/himself as paparazzi, chasing learners through the streets; the agent may also be a groundhog in the Loess Plateau, a Bregalad in the Smoky Forest, or a fortuneteller in a fortune-telling house (see Table 1). They interview the learners and initiate reflection-promoting conversations.
- The “How” principle: To ensure that the interactions are as conversational as a typical daily chat, AI technologies (semantic networks and natural language processing) were used to control the conversational context (as described in the later section). In this way, the conversations between agents and learners appear natural and realistic.
- The “What” principle: Five of the eight intelligences in the MI scheme (specifically, the linguistic, logical-mathematical, spatial, interpersonal, and intrapersonal intelligences) were used as the basis for questions that prompted self-reflections. All of the questions asked by the agents are generated in accord with the simulation scenarios/activities that the learners have just explored. Because the tasks in the simulation environment provide ample opportunities for hands-on experience and practice that would enable learners to use various abilities, the

appearances of the reflection-prompting agents are to prompt learners to reflect on their intelligence preferences.

- The “When” principle: To facilitate real-time self-reflection, the reflections are immediately elicited by the reflection-prompting agents while the learners work on relevant simulation tasks. There is no interval between reflection and learning because reflection is part of the learning process.
- The “Where” principle: Our goal was to maintain continuity between the teaching/learning and the reflection activities. Table 1 indicates the places and activity themes: (a) the astrology club (fortune-telling house) where the astrologers (fortunetellers) talk with the learners and predict their future life events; (b) several cities along a train tour where the paparazzi are interested in the learners’ tasks and then collect learners’ information for a news report; and (c) the monster dens where learners exchange information with gatekeeper monsters for the right to enter other dens and seek to perform more tasks. The reflections do not interrupt the progress of the simulation scenarios.

*Table 1.* The activity themes, scenes, and roles of the reflection-prompting agents.

Reflection Activity Themes	Scenes	Agents’ Roles
<b>Fortune-telling</b> Astrologers (fortunetellers) ask users questions and make predictions for what they will experience next week in terms of close relationships, friendships, schoolwork, financial budgets, and work/career.	<b>Two booths:</b> Astrology club Fortune-telling house	<b>Astrologers:</b> The Astrology Prince, Vivian, and Liz Tang. <b>Fortunetellers:</b> Miss Jen-Yi Lin and Miss Yu Yang.
<b>One-day Train Tour</b> Users travel by train to several cities. Paparazzi follow the users and conduct interviews along the tour.	<b>Three day tours:</b> One-day tour to Hualien One-day tour to Taipei One-day tour to Kaohsiung	<b>Paparazzi:</b> for example, Stalkerazzi
<b>Treasure Hunt in Monster Dens</b> Users are within a labyrinth and must walk through several gateways guarded by monsters when they need to move from the first to the second floor. By answering questions, users earn keys or tools to enter the second floor.	<b>Five settings:</b> Mountain of Doom Gold Cave of Darkness Loess Plateau Dead Marshes Smoky Forest	<b>Monster:</b> Balrog <b>Monster:</b> Gold Python <b>Toy:</b> Groundhog <b>Monster:</b> Siren <b>Monster:</b> Bregalad

### The communication methods between the reflection-prompting agents and learners

We edited the questions using semantic network technology. The question ordering follows semantic network links (an example is illustrated in Figure 3).

The reflection-prompting agents would provide appropriate feedback depending on whether the learner’s replies are positive or negative (as reported below), thereby affecting the subsequent question. The question ordering is different for each student. To ensure that each student is asked each question only once, the design of the semantic networks is kept simple. A semantic network comprises only 5-10 questions. Our fifty self-reflection questions were divided into several semantic networks, integrated with the relevant scripts and scenes, and posted by the agents at the appropriate time. Several semantic networks aligned with potential script developments. Using semantic networks, we are able to develop dialogue lines for agents’ social conversations.

We analyzed the learners’ answers using natural language processing technology to identify the wording in sentences and classify the answers into three categories: positive answers, negative answers and neutral comments. Neutral comments (e.g., “not bad” and “so-so”) receive zero points, or answers implying uncertain or nonsensical ideas trigger an agent to repeat the question. Other sentences are divided into adverbs, negatives and adjectives/verbs (see Table 2). If the replies include an adverb of degree (e.g., very, more, much, rather, quite, or too), the adverb section will receive two points; if no adverb is found, it will receive one point. If the replies include a negative, the negative section will receive minus-one point; if no negative is found, it will receive one point. If the replies include a positive adjective or verb, the adjective/verb section will receive one point; if a negative adjective or verb is identified, it will receive minus-one point. Moreover, we found that high school students frequently used slang and expression symbols when answering the questions in the pilot test. For example, “I damn like this...”, “I super hate...”, and “XD”. Therefore, to enhance the identification ability of the natural language processing technology, we expanded

the conversation database to include popular slang and expression symbols. Slang for expressing degree (e.g., damn and super) is regarded as an adverb of degree (receiving two points). Positive expression symbols (e.g., ^^, :), and XD, meaning smile or laugh) are regarded as positive adjectives/verbs (receiving one point); negative expression symbols (e.g., ><, :(, and Q\_Q, meaning anger or cry) are regarded as negative adjectives/verbs (receiving minus-one point). Points for individual words are multiplied to produce a total score (from -2 to +2) for each sentence/comment. Several examples are demonstrated below.

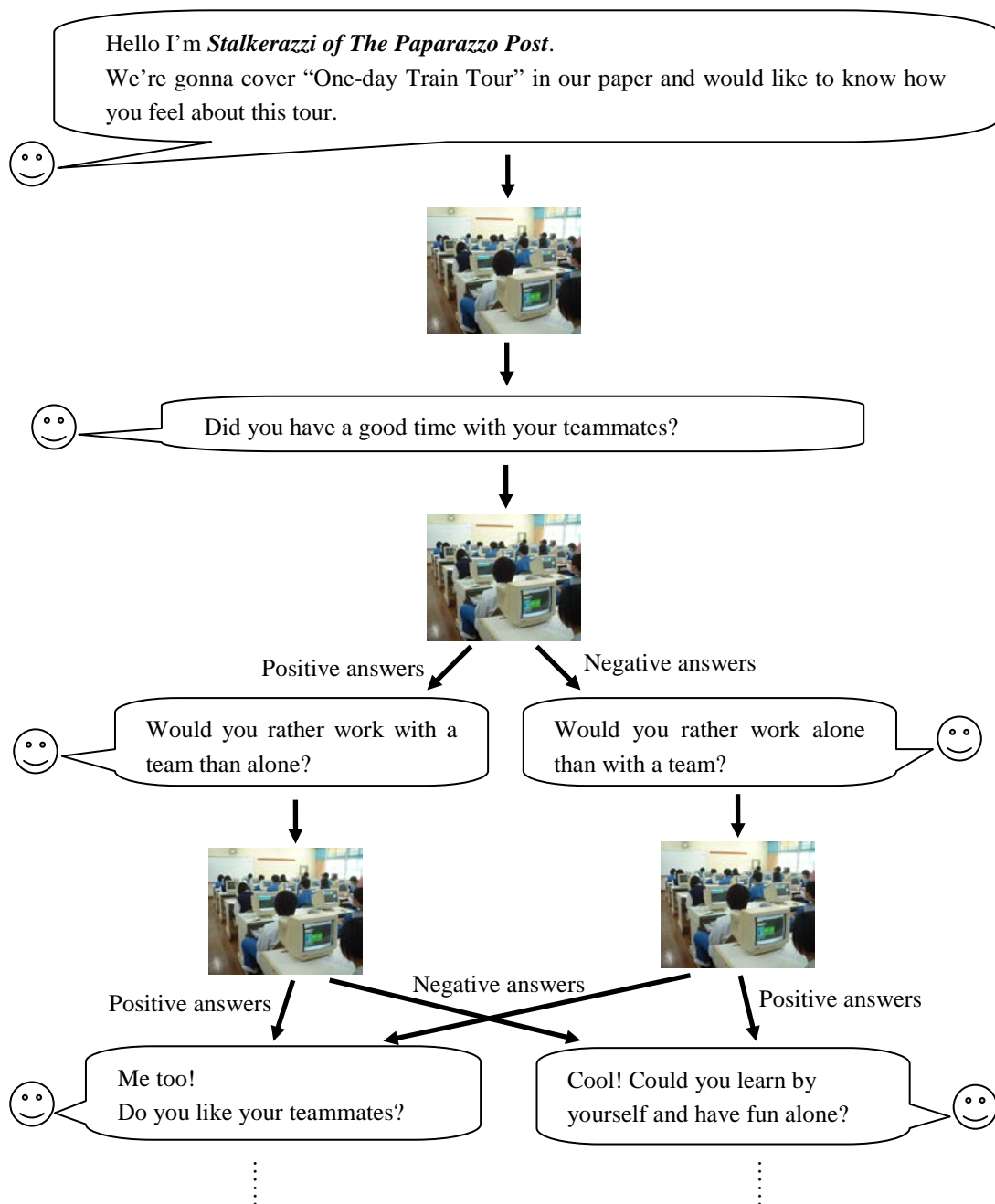


Figure 3. A partial example of a semantic network.



I love him very much.	Neg=1, V=1, Adv=2 → 1*1*2=2 → +2 points
I damn like swimming.	Neg=1, V=1, Adv=2 → 1*1*2=2 → +2 points
Yes, I like it.	Neg=1, V=1, Adv=1 → 1*1*1=1 → +1 point
It is not bad.	→ 0 points
I don't like to play alone.	Neg=-1, V=1, Adv=1 → -1*1*1=-1 → -1 point
><	Neg=1, V=-1, Adv=1 → 1*-1*1=-1 → -1 point
I really hate to keep a diary.	Neg=1, V=-1, Adv=2 → 1*-1*2=-2 → -2 points
I'm not sure.	→ Repeat the question

If the total score is smaller than 0, the sentence is considered to be a negative answer; otherwise, it is a positive answer. The primary purpose of this scoring system is to make it possible to conduct further quantitative analyses and to make comparisons between MI self-reflections collected through the agent interviews and those obtained using the paper-pencil questionnaires.

Table 2. The words are used to analyze learner comments.

Adverbs	Negatives	Positive Adjectives/verbs	Neutral comments	Uncertain or nonsensical ideas
very, too, greatly, pretty, more, much, extremely, often, damn, always, super, the most, fairly, quite, really, absolutely...etc.	not, none, without...etc.  <b>【-1 point】</b>	like, love, yes, fit in with, may, ok, clear, can, agree with, proper, able, sure, want, ^^, :), XD...etc.  <b>【+ 1 point】</b>	just on, not bad, so-so, no affect, acceptable, passable, common...etc.  <b>【0 points】</b>	I don't know, uncertain, not sure...etc, or nonsense  <b>【Ask again】</b>
<b>【+2 points】</b>				
No adverb	No negative	Negative Adjectives/verbs		
<b>【+ 1 point】</b>	<b>【+ 1 point】</b>	bad, dislike, hate, no, less, disagree, ><, :(, Q_Q...etc  <b>【-1 point】</b>		

## Methods

### Participants

The participants were 117 tenth-grade students from the Information Processing Program at a vocational high school in Taiwan. According to prior performance records from the national senior high school entrance examination, students in this school were ranked in the top 30% of all senior high school students in Taiwan.

### Measures

Because the participants were high school students in Taiwan, we used Tai's (2002) Chinese High School Students Inventory of Multiple Intelligences, which was developed to measure high school students' MI preferences. The 80-item questionnaire was modified from Armstrong's (1994) MI Inventory for Adults. As mentioned previously, our focus was on five of the eight categories of intelligence; thus, the study used 50 of the original questionnaire items for our purposes, with 10 items relating to each of the following intelligence categories: linguistic, logical-mathematical, spatial, interpersonal, and intrapersonal. All of the items are measured on a five-point Likert scale. Item analysis was used to assess the quality of individual questions and a combination of reliability, and factor

analyses were used to evaluate the overall reliability and construct validity. The Cronbach's alpha coefficients, which ranged from 0.8015 to 0.9102, were regarded as acceptable.

## Procedures

The test was conducted in computer classrooms over a three-week period. Before the simulation activities occurred, the students were required to fill in a pre-simulation paper-pencil questionnaire (20 minutes), read through the instructions for operating the simulation system (60 minutes), and then complete the pre-simulation agent interview, Fortune-telling (30 minutes, no learning tasks, only simulated interview situations in which agents act as the astrologers), during week 1. The first simulation activity, One-day Train Tour (120 minutes, including the agent interview, agents act as the paparazzi), was completed during week 2. The second simulation activity, Treasure Hunt in Monster Dens (120 minutes, including the agent interview, agents act as the monsters), and the post-simulation paper-pencil questionnaire (20 minutes, it was filled after all of the simulation activities) were conducted during week 3.

## Results

### Two thirds of the participants successfully completed the simulation activities and agent-mediated self-reflections

We examined each participant's log and record and found that of the 117 high school students, five students could not finish the first simulation activity and 45 students could not finish the second simulation activity (including the agent-prompted reflections in both cases) within the preplanned 120-minute time window. Because the second simulation activity is more difficult than the first one, several students could not generate solutions to those tasks in time. We observed that the time needed by each learner to solve the problems was different, and therefore we suggest that the time limitation or difficulty level should be adjusted to raise the success rate for that activity.

The following is a real conversation transcribed from our system logs. The agent named Golden Python, a virtual monster, interacted with Wantai, a participant.

Golden Python: Hello, I am Golden Python. I am extremely wealthy with unlimited gold, and I love dispensing gold products to people. If you want to receive my gifts, you'll have to be friends with me.  
Can we be friends?

Wantai: Yes.

Golden Python: Awesome! I'm so excited!  
Besides me, do you have other good friends?

Wantai: Yes. I do have some.

Golden Python: Oh... I envy you.  
Do you like outdoor activities?

Wantai: I love outdoor activities very much!

Golden Python: Well then, let's go do something outdoors together next week!  
Besides that, do you have any other hobbies?

Wantai: Yes. I like to collect badges.

Golden Python: Hm...I see.  
Do you like painting?

Wantai: Yes. I do.

Golden python: Excellent! I love painting too!  
Do you like entertainment, like TV programs, movies or magic shows?

Wantai: Yes. I like to watch TV and movies.

Golden python: Wow! Your life is really interesting and exciting!  
Do you want to know more about me?

Wantai: Yes! I would like to.

Golden python: I live in this massive gold mine. I have experienced vicissitudes and feel indifferent toward all these people wasting their lives striving for power, fame and wealth. People trick each other and constantly fight their peers. I feel very lonely!

Do you enjoy being with a crowd of people?  
Wantai: Yes, I like to be with friends.  
Golden python: Me too.  
Do you like to be the leader in a group of people?  
Wantai: So-so.  
Golden python: Huh?  
Do you like to play toy models?  
Wantai: Yes. I like that.  
Golden python: OK. I have to go! It is my pleasure to meet you! I'd like to give you a "Golden Toy Model" in honor of our meeting!

### Participants favored the agent-prompted reflection over a paper-pencil questionnaire

The agent interview was preferred by 72.03% of all the participants, whereas 3.50% of the participants preferred the questionnaire. The remaining 24.47% perceived no difference between the two. To further analyze the rationale for each choice, we summarized and categorized participant rationales into the following points.

The reasons for preferring the agent interview include the following rationales:

- It was interactive, interesting, funny and laid-back, just like chatting. I was not bored and felt relaxed, cheery, and amused.
- I answered more carefully because I received some feedback that gave me suggestions.
- It was more flexible, and I could elaborate upon my ideas, thoughts and feelings.
- It was a virtual scenario that accurately simulated a real interview. I found it to be very realistic. Sometimes, the system would ask me to repeat when it could not read my words.
- It was more efficient and interesting to test and play a game at the same time. It could arouse my interest and help me feel more committed to these activities.
- Sitting in front of a computer is quite unique. I was tired of the paper-and-pencil questionnaire.

The reasons for preferring the questionnaire include the following rationales:

- I did not have to read sentences that were irrelevant to the question.
- I did not have to type. The computer sometimes could not read what I meant.
- I had enough time to answer carefully when I used the questionnaire; I did not have sufficient time to think through my answers during the simulation course because the questions kept popping up.

The primary reason for choosing "no difference" follows:

- Both formats were designed to answer questions.

The results indicate that a great majority of students preferred the agent interview to the questionnaire. As a result of the interactions, feedback, and scenarios in the simulation environment, students generally felt that the agent interview was interesting, entertaining, lively, and novel.

### The reflections collected by the agent interviews were in accord with those collected through the paper-pencil questionnaires

Table 3 lists the self-reflections (means and standard deviations for MI subscales) (a) collected through the paper-pencil questionnaires on two occasions: pre-simulation and post-simulation (N = 117), and (b) collected through the agent interviews on three occasions: pre-simulation (N = 117), during the first simulation activity (N = 112) and during the second simulation activity (N = 72).

*Table 3.* The mean and standard deviation statistics for five MI reflections collected on two occasions (pre- and post-simulation) and through two formats (paper-pencil questionnaire and agent interview).

Occasions	Pre-Simulation		Post-Simulation		
	Paper-pencil	Agent Interview (N=117)	During 1 <sup>st</sup> sim	During 2 <sup>nd</sup> sim	Paper-pencil
Formats			Agent Interview	Agent Interview	

	(N=117)		(N=112)		(N=72)		(N=117)	
MI's	M	SD	M	SD	M	SD	M	SD
Linguistic	3.11	0.57	3.16	0.37	3.29	0.44	3.41	0.41
Logical	3.01	0.62	3.16	0.45	3.27	0.42	3.44	0.41
Spatial	3.49	0.61	3.48	0.38	3.58	0.31	3.74	0.30
Interpersonal	3.71	0.65	3.75	0.32	3.77	0.26	3.82	0.28
Intrapersonal	3.40	0.50	3.43	0.36	3.54	0.27	3.64	0.28

Correlation analyses were conducted to determine whether the reflections about the same MI's collected through hetero-occasion same-format means as well as the reflections collected through hetero-format same-occasion means were correlated (in Table 4). The correlation (of the same trait) of hetero-occasion same-format is an index of stability that should be significant and high in magnitude, representing the fact that the two datasets belong to the same theoretical construct. The variance primarily derives from the changes in occasions (in this study, before and after the simulation exploration and problem solving). The results show that the correlations between the five MI reflections collected through the paper-pencil questionnaires on pre- and post-simulation occasions were all significant with high magnitudes ( $r_s = .538 \sim .831, p < .01$ ). The correlations of the five MI reflections collected through the agent interviews on different occasions were also significant: between the pre-simulation and the 1st simulation reflection collections,  $r_s = .570 \sim .618 (p < .01)$ ; between the pre-simulation and the 2nd simulation reflection collections,  $r_s = .374 \sim .645 (p < .01)$ ; and between the 1st simulation and the 2nd simulation reflection collections,  $r_s = .519 \sim .672 (p < .01)$ . The correlations of reflections collected by hetero-format hetero-occasion means were either non-significant (e.g.,  $r = -.065$ ) or significant but low in magnitude (e.g.,  $r = .208, p < .05$ ).

The correlation (same-trait) collected through hetero-format same-occasion means constitutes an index known as convergent validity in psychometrics (MTMM, Crocker & Algina, 1986). The very same trait evaluated by two different methods should be significant but to a lesser degree than the stability index mentioned above. Under this type of comparison, the datasets share the same theoretical concept; thus, the major variance derives from various measurement methods (also called the method effect). The results listed in Table 4 indicate that for the pre-simulation occasion, the same MI reflections collected through hetero-formats (paper-pencil questionnaire vs. agent interview) were significantly correlated ( $r_s = .411 \sim .638, p < .01$ ). Similarly, for the post-simulation occasion, the same MI reflections collected through hetero-formats (paper-pencil questionnaire vs. agent interview in the 2nd simulation) were also correlated ( $r_s = .314 \sim .485, p < .01$ ). A comparison of the correlations obtained if only the individuals present in all collections were considered demonstrates that the same-format hetero-occasion correlations (stability) are higher than the hetero-format same-occasion correlations (convergent validity).

Table 4. Correlations between hetero-format data (paper-pencil questionnaire and agent interview) and hetero-occasion data (pre- and post-simulation activities).

Formats Occasions	Hetero-Format Same-Occasion			Same-Format Hetero-Occasion		
	P / A Pre-simulations (N=117)	P / A Post-simulations (N=72)	A Pre/1 <sup>st</sup> sim (N=112)	A Pre/2 <sup>nd</sup> sim (N=72)	A 1 <sup>st</sup> / 2 <sup>nd</sup> sims (N=72)	P Pre/Post (N=117)
MI's						
Linguistic	.445**	.340**	.570**	.645**	.672**	.742**
Logical	.638**	.314**	.618**	.496**	.519**	.831**
Spatial	.457**	.485**	.578**	.374**	.559**	.689**
Interpersonal	.411**	.334**	.594**	.565**	.618**	.538**
Intrapersonal	.483**	.387**	.592**	.439**	.641**	.684**

Note. P = paper-pencil, A = Agent interview; \*\*  $p < .01$

As discussed above, statistics demonstrated that the MI preference self-reflections collected through the agent interviews were adequately in accord with the self-reflections collected through the traditional paper-pencil questionnaires. Both self-reflection formats are capable of evaluating self-reflections regarding the five intelligence categories. Thus, it is feasible and appropriate to use the agent interview method to collect and transform conversational comments into quantitative scores.

## Discussions

This study developed a new method for collecting self-reflections via agent interviews in a simulation environment. The results support the notion that this new method is reliable and convergent with the traditional paper-pencil methodology. Furthermore, the MI reflections collected through the agent interviews remained rather stable across different occasions. Thus, self-reflection can be effectively prompted not only through questionnaires but also through agent interviews in simulation environments.

Our design allows learners to reflect on themselves through fun, vivid and attractive tasks, thereby enhancing students' motivation and interest in self-reflection. The features of the simulation technology, such as immediacy of response, free options, and bi-directional communications, help enhance the interactions between the users and the medium. Furthermore, educational agents enrich the social context in a learning environment by providing virtual participants that enhance the member multiplicity and by supporting means of fostering communication with real participants. Our reflection collections are blended with the interesting characters, story situations, and scenes in a simulation environment. This method elicits students' curiosity. The participants reported that they preferred casual conversations with the agents to filling in questionnaires.

Investigators have long recognized the need for an assessment tool that is more representative of life experiences than laboratory assessments, global questionnaires, or observer ratings (Redelmeier & Kahneman, 1996). Our design can collect data in a simulation environment in which learners respond to questions throughout the course of working on tasks in a setting that is very similar to their natural learning environment. The advantage of using agent interviews in simulation is that the self-reflections are collected while the participants are still exploring or have just solved problems. The self-reflections are collected in real-time while the problem-solving situation remains ongoing; it may reduce situational biases. Furthermore, the educational value of computer-assisted assessment arises largely from its capacity to provide students with immediate feedback on their comprehension level. Agent interviews can provide substantial and rich stimuli for a learner to display an index of instantaneous reflections about his/her own cognition, comprehension, abilities and affective states during the learning process. The participants do not need to reflect their competences/preferences based upon retrospective memories that may produce memory biases.

## Conclusions

In this study, we introduce a new practical application of simulation as an effective learning tool: simulations can prompt multi-faceted self-reflections and thereby enhance self-awareness. We suggest that self-awareness about personal competence, values and affections is critical for high school students in choosing an appropriate undergraduate major program. The introduction of simulations that prompt self-reflections thus provides not only short-term implications for classroom learning but also long-term prospects for career decision guidance.

Simulations meet various teaching requirements and have been widely applied in a broad range of educational and training courses. To understand students' learning processes, performance, or outcomes in simulated environments, three types of measurements are generally applied. First, some systems might automatically record learner outcomes, such as parameter changes (critical indicators presented in the simulation process) and user behavior timestamps (Shute, Ventura, Bauer, & Zapata-Rivera, 2009; Tan et al., 2010). These outcome records consist mostly of quantitative summaries presented after the completion of simulation activities. Quantitative data provide unambiguous feedback for teachers. However, if learners do not learn well, teachers cannot retrieve step-by-step detailed processes of where the students have failed. Second, some systems automatically record detailed action transformations during simulation activities, such as group work interactions or step-by-step decision-making processes (McLaughlan & Kirkpatrick, 2004; Tan et al., 2010). The information generated through this process is often tremendous volume and therefore can make it difficult to formulate a summary report. Teachers also lack the analytical tools to diagnose learning processes, and therefore this type of system is not practical for improving teaching. Finally, some teachers use oral and written briefings, debriefing sessions, interviews, or questionnaires to collect learners' reflections on a simulated learning course (Asakawa & Gilbert, 2003; Martin, 2006). These reflections can help teachers grasp positive or negative impact factors during a learning course. Nonetheless, when briefings, interviews, or questionnaires are conducted simultaneously with the simulation course, they may interrupt immersed learning and reduce the quality of simulation-based learning. If these responses are not measured during the simulation course, some situational and memory biases may occur. Our design retains the advantages of the three

types of measurements reported above and mitigates their disadvantages. This self-reflection design provides an adequate method for collecting students' reflections on their own cognition and affective states during the learning process in a simulated environment and assists teachers in understanding students' learning situations in simulation-based e-learning.

We acknowledge at least two study limitations and suggest future studies. First, because this design is an innovative attempt to integrate self-reflection with simulation, it is impossible to test the main effect of this method using typical experimental methods, which would require a control group that experiences traditional classroom instruction. This study used observational techniques to collect students' self-reflections based on the MI scheme; the evidence about its effect in supporting conceptual learning or self-regulated learning is thus clearly preliminary. Future studies are encouraged to adopt mixed research methods (Maxwell & Loomis, 2003) to obtain a deeper understanding of the effects created when teachers introduce innovative assessment techniques in simulations. Second, we used a text-based online virtual reality system as the learning environment; however, this type of virtual reality system is not popular at the present time. As graphics-based virtual reality systems are popular and available, future studies could adopt these systems that incorporate agent interview mechanisms to prompt self-reflection.

## Acknowledgements

This study is financially supported by National Science Council “NSC 99-2511-S-009-011-MY3,” “NSC 100-2631-S-009-001,” and “NSC 100-2511-S-126-006-MY2.”

## References

- Armstrong, T. (1994). Multiple intelligences: Seven ways to approach curriculum. *Educational Leadership*, 53(3), 26–28.
- Asakawa, T., & Gilbert, N. (2003). Synthesizing experiences: Lessons to be learned from Internet-mediated simulation games. *Simulation & Gaming*, 34(1), 10–22.
- Atkinson, R. K., Mayer, R. E., & Merrill, M. M. (2005). Fostering social agency in multimedia learning: Examining the impact of an animated agent's voice. *Contemporary Educational Psychology*, 30(1), 117–139.
- Bank, L., Dishion, T. J., Skinner, M., & Patterson, G. R. (1990). Method variance in structural equation modeling: Living with “glop”. In Patterson, G. R. (Ed.), *Depression and Aggression in family interaction: Advances in family research* (pp. 247–279). Hillsdale, NJ: Erlbaum.
- Carbonell, J. R. (1970). AI in CAI: An artificial-intelligence approach to computer-assisted instruction. *IEEE Transactions on Man-Machine Systems*, 11(4), 190–202.
- Chen, C.-H., & Howard, B. (2010). Effect of live simulation on middle school students' attitudes and learning toward science. *Educational Technology & Society*, 13(1), 133–139.
- Chen, Y.-L., Hong, Y.-R., Sung, Y.-T., & Chang, K.-E. (2011). Efficacy of simulation-based learning of electronics using visualization and manipulation. *Educational Technology & Society*, 14(2), 269–277.
- Chou, C.-Y., Chan, T.-W., & Lin, C.-J. (2003). Redefining the learning companion: The past, present, and future of educational agents. *Computers & Education*, 40(3), 255–269.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando, FL: Harcourt Brace Jovanovich College.
- Cuenca López, J. M., & Martín Cáceres, M. J. (2010). Virtual games in social science education. *Computers & Education*, 55(3), 1336–1345.
- de Bruin, A. B. H., Rikers, R. M. J. P., & Schmidt, H. G. (2005). Monitoring accuracy and self-regulation when learning to play a chess endgame. *Applied Cognitive Psychology*, 19(2), 167–181.
- Eskrootchi, R., & Oskrochi, G. R. (2010). A study of the efficacy of project-based learning integrated with computer-based simulation - STELLA. *Educational Technology & Society*, 13(1), 236–245.
- Gardner, H. (1983). *Frames of mind: The theory of Multiple Intelligence*. New York, NY: Basic Books.

- Gardner, H. (1993). *Multiple intelligences: The theory in practices*. New York, NY: Basic Books.
- Gardner, H. (1998). A multiplicity of intelligences. *Scientific American Presents*, 9(4), 18–23.
- Gibbons, M. (2002). *The self-directed learning handbook: Challenging adolescent students to excel*. San Francisco, CA: Jossey-Bass.
- Greene, J., & Azevedo, R. (2007). A theoretical review of Winne and Hadwin's model of self-regulated learning: New perspectives and directions. *Review of Educational Research*, 77(3), 334–372.
- Hsieh, C.-H., & Sun, C.-T. (2007). Robots in educational MUDs: Instructional roles and functions. *International Journal of Instructional Media*, 34(1), 83–90.
- Huang, C.-C., Yeh, T.-K., Li, T.-Y., & Chang, C.-Y. (2010). The idea storming cube: Evaluating the effects of using game and computer agent to support divergent thinking. *Educational Technology & Society*, 13(4), 180–191.
- Ifenthaler, D. (2012). Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios. *Educational Technology & Society*, 15(1), 38–52.
- John, N. W. (2007). The impact of Web3D technologies on medical education and training. *Computers & Education*, 49(1), 19–31.
- Lin, H., & Sun, C.-T. (2003). Problems in simulating social reality: Observations on a MUD construction. *Simulation & Gaming*, 34(1), 69–88.
- Liu, T.-C. (2010). Developing simulation-based computer assisted learning to correct students' statistical misconceptions based on cognitive conflict theory, using "Correlation" as an example. *Educational Technology & Society*, 13(2), 180–192.
- Martin, A. (2003). Adding value to simulation/games through Internet mediation: The medium and the message. *Simulation & Gaming*, 34(1), 23–38.
- Mauldin, M. L. (1994). Chatterbots, tinymuds, and the turing test: Entering the loebner prize competition. *Proceedings of the Twelfth National Conference on Artificial Intelligence*. Retrieved from <http://www.aaai.org/Papers/AAAI/1994/AAAI94-003.pdf>
- Maxwell, J. A., & Loomis, D. M. (2003). Mixed methods design: An alternative approach. In Tashakkori, A. & Teddlie, C. (Eds.) *Hand-book of mixed methods in social and behavioral research* (pp. 241–272). Thousand Oaks, CA: Sage.
- McLaughlan, R. G., & Kirkpatrick, D. (2004). Online role play: Design for active learning. *European Journal of Engineering Education*, 29(4), 477–490.
- Meij, J. van der, & de Jong, T. (2006). Supporting students' learning with multiple representations in a dynamic simulation-based learning environment. *Learning & Instruction*, 16(3), 199–212.
- Ministry of Education (2012). *Program description for "Program for promoting teaching excellence of universities."* Retrieved August 29, 2012, from [http://www.csal.fcu.edu.tw/Edu/program\\_start.asp](http://www.csal.fcu.edu.tw/Edu/program_start.asp)
- Moreno, R., & Mayer, R. E. (2005). Role of guidance, reflection, and interactivity in an agent-based multimedia game. *Journal of Educational Psychology*, 97(1), 117–128.
- Padiotis, I., & Mikropoulos, T. A. (2010). Using SOLO to evaluate an educational virtual environment in a technology education setting. *Educational Technology & Society*, 13(3), 233–245.
- Peterson, M. (2010). Computerized games and simulations in computer-assisted language learning: A meta-analysis of research. *Simulation & Gaming*, 41(1), 72–93.
- Ranalli, J. (2008). Learning English with the Sims: Exploiting authentic computer simulation games for L2 learning. *Computer Assisted Language Learning*, 21(5), 441–455.
- Redelmeier, D. A., & Kahneman, D. (1996). Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. *Pain*, 66(1), 3–8.
- Rosen, K. R. (2008). The history of medical simulation. *Journal of Critical Care*, 23(2), 157–166.
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, 54(2), 93–105.
- Schwarz, N. (2007). Retrospective and concurrent self-reports: The rationale for real-time data capture. In A. A. Stone, S. S. Shiffman, A. Atienza, & L. Nebeling (Eds.), *The Science of Real-Time Data Capture: Self-Reports in Health Research* (pp. 11–26). New York, NY: Oxford University Press.

Schwarz, N., & Sudman, S. (1994). *Autobiographical memory and the validity of retrospective reports*. New York, NY: Springer Verlag.

Shute, V. J., Ventura, M., Bauer, M. I., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In Ritterfeld, U., Cody, M., & Vorderer, P. (Eds.), *Serious games: Mechanisms and effects* (pp. 295–321). Mahwah, NJ: Routledge.

Stanford, P. (2003). Multiple Intelligence for every classroom. *Intervention in School and Clinic*, 39(2), 80–85.

Tai, C.-Y. (2002). *Application of a multiple intelligences model in English instruction in junior high school: Implication of an alternative English instruction model in the nine-year education program*. (Unpublished master's thesis). National Kaohsiung Normal University, Kaohsiung, Taiwan.

Tan, K.-H., Tse, Y.-K., & Chung, P.-L. (2010). A plug and play pathway approach for operations management games development. *Computers & Education*, 55(1), 109–117.

Wang, Q. (2001). Culture effects on adults' earliest childhood recollection and self-description: Implications for the relation between memory and the self. *Journal of Personality and Social Psychology*, 81(2), 220–233.

Martin, A. (2006). Adding value to simulation/games through Internet mediation: The medium and the message. *Simulation Gaming*, 34(1), 23–38.

Wu, L., & Looi, C.-K. (2012). Agent prompts: Scaffolding for productive reflection in an intelligent learning environment. *Educational Technology & Society*, 15(1), 339–353.



## An Evaluation of the Learning Effectiveness of Concept Map-Based Science Book Reading via Mobile Devices

**Che-Ching Yang<sup>1</sup>, Gwo-Jen Hwang<sup>2\*</sup>, Chun-Ming Hung<sup>3</sup> and Shian-Shyong Tseng<sup>4</sup>**

<sup>1</sup>Department of Computer Science, National Chiao Tung University, 1001 Ta Hsueh Road, Hsinchu, 300, Taiwan //

<sup>2</sup>Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology 43, Sec.4, Keelung Rd., Taipei, 106, Taiwan // <sup>3</sup>Department of Information and Learning Technology, National University of Tainan No. 33, Sec. 2, Shulin St., Tainan city 70005, Taiwan // <sup>4</sup>Department of Applied Informatics and Multimedia, Asia University 500 Lioufeng Road, Wufeng, Taichung, 413, Taiwan // jerome@cis.nctu.edu.tw // gjhwang.academic@gmail.com // hcm@live.htps.tn.edu.tw // sstseng@asia.edu.tw

\*Corresponding author

(Submitted May 09, 2012; Revised October 15, 2012; Accepted October 20, 2012)

### ABSTRACT

Printed books have long been an important means for people to obtain knowledge; nevertheless, educators have pointed out problems of learning from some printed books, in particular, the lack of learning supports, such as supplementary materials and learning guiding tools. In this study, a concept map-oriented ubiquitous learning approach is proposed for supporting printed science book reading activities for children via mobile devices with QR-code facilities. A quasi-experiment has been conducted in an elementary school to evaluate the effectiveness of the proposed approach, in which 92 sixth graders participated. The experimental results show that the concept map-oriented ubiquitous learning approach is significantly more helpful to the students in reading printed books than traditional book reading and the conventional ubiquitous learning approach in terms of learning achievements; moreover, the students had a high level of acceptance of such a mobile technology-assisted learning system in terms of “ease of use”, “usefulness” and “attitude and intention of future use.”

### Keywords

Concept maps, Ubiquitous learning, Printed book reading, QR codes

### Introduction

Printed books have been advocated as an important learning resource (Fletcher & Reese, 2005; Ortiz, Stowe, & Arnold, 2001) and have been linked to language development (Whitehurst et al., 1988), emergent literacy (Allor & McCathren, 2003; Whitehurst & Lonigan, 1998), oral and written skills (Whitehurst & Lonigan, 1998), and even science learning (Chen, Teng, Lee, & Kinshuk, 2011). When reading books, children activate and integrate their prior knowledge with the book content, use strategic processes to identify key concepts, synthesize and summarize information, make inferences, and participate in the story (Paris & Paris, 2003). However, for learning content that is abstract, such as science or mathematics, children might lose interest in reading the books without learning supports (Mantzicopoulos & Patrick, 2010, 2011).

In the past decades, owing to the advancement and growing popularity of computer and network technologies, a considerable number of digital books have been developed to provide a more convenient channel for people to learn; however, numerous learners still prefer reading printed books (Ackerman & Goldsmith, 2008; Chao & Chen, 2009; Chen, Teng, Lee, & Kinshuk, 2011; O'Hara & Sellen, 1997). Furthermore, researchers have reported the benefits of learning with printed books; for example, O'Hara and Sellen (1997) pointed out the major advantage of printed book reading, including (1) free-text annotation, which helps learners highlight, underline, summarize, and/or annotate; (2) quick navigation, which helps learners navigate the text quickly and automatically; and (3) flexibility of spatial layout, which helps learners gain a sense of overall structure, cross referencing and interleaving of reading/writing. The study of Ackerman and Goldsmith (2008) showed that the learning achievement of students who learned with printed texts was better than that of students who learned with digital texts. That is, printed books remain an effective means of learning. On the other hand, researchers have pointed out some problems of reading printed books, including the lack of learning supports, such as supplementary materials and learning guiding tools (Dünser & Hornecker, 2007). Therefore, it remains an important issue to provide learning supports for printed book reading (Chen, Teng, Lee, & Kinshuk, 2011).

Recently, the advancement of mobile and wireless communication technologies has further offered an opportunity for providing support for reading activities with digitalized supplementary materials. For example, several studies have demonstrated the convenience of using mobile devices in supporting and augmenting paper-based learning (Chao & Chen, 2009; Chen, et al., 2011; Ozcelik & Acarturk, 2011). Hwang, Tsai and Yang (2008) further defined the learning approach that employs mobile, wireless communication and sensing technologies to enable students to learn in real-world environments with access to digital resources as *context-aware ubiquitous learning* (u-learning).

On the other hand, scholars have emphasized the importance of providing scaffolding, a kind of instructional support, to help learners achieve goals that they cannot accomplish on their own (Kuo, Hwang, & Lee, 2012; Peng et al., 2009; Wood, Bruner, & Ross, 1976). For example, the studies of Chu, Hwang and Tsai (2010) and Hwang, Chu, Lin and Tsai (2011) showed the effectiveness of using knowledge construction tools in helping students conduct ecology observations for a natural science course. Chen et al. (2011) further indicated that the provision of scaffolding could benefit learners in reading. Among various learning strategies or tools, concept maps have been considered as being an effective tool for facilitating meaningful learning by helping students to link their prior knowledge with new experiences as well as organizing the accumulated knowledge (Liu, Chen, & Chang, 2010).

In this study, we augment paper-based reading activities in an elementary school natural science class by developing a reading system using smartphones in association with QR (Quick Response) code technology to provide students with a way of linking printed science books with corresponding digital materials and concept maps. Moreover, an experiment has been conducted to demonstrate the effectiveness of the proposed approach.

## Literature review

### Computer and mobile technologies in reading activities

The popularity of computer technologies has provided opportunities to conduct various kinds of learning activities. Earlier studies mainly employed multimedia technology to help students improve their reading comprehension (Garza, 1991; Hoven, 1999). Later, some researchers employed more advanced computer technologies, such as Virtual Reality (VR) or Augmented Reality (AR), to assist children in reading printed books by providing three-dimensional virtual content (McKenzie & Darnell, 2003; Yang et al., 2009).

The advancement of portable devices (e.g., PDAs, smartphones and e-books) and wireless communication technology has brought the learning technology into a new era, in which the potential of mobile and ubiquitous learning or one-to-one learning has been noted by researchers (Chan et al., 2006; Hwang, Tsai, & Yang, 2008). The benefit of using mobile and wireless communication technologies to provide learners with learning supports has been reported by many researchers, and includes learning guidance for in-field activities (Chiou, Tseng, Hwang, & Heller, 2010; Chu, Hwang, & Tseng, 2010; Hwang, Tsai, Chu, Kinshuk, & Chen, 2012) or in-class activities (Chang, Lan, Chang, & Sung, 2010; Wong, Chin, Tan, & Liu, 2010), as well as the provision of help-seeking (Yang, 2006) and assessment (Coulby, Hennessey, Davies, & Fuller, 2011; Hwang & Chang, 2011) mechanisms. Those learning systems have been applied to various practical applications, such as science learning (Hwang, Chu, Shih, Huang, & Tsai, 2010; Liu, Peng, Wu, & Lin, 2009; Rogers, Price, Randell, Fraser, Weal, & Fitzpatrick, 2005; Shih, Chuang, & Hwang, 2010), social science learning (Hwang & Chang, 2011) and language courses (Looi et al., 2010, 2009; Ogata, Matsuka, El-Bishouty, & Yano, 2009).

Chan et al. (2006) and Roschelle and Pea (2002) have indicated that such one-to-one learning has augmented physical space and leveraged topological space. They pointed out that the affordance of one-to-one learning using handheld devices is the latest trend in learning, whereby digital augmentation might break down many of the physical barriers to learning.

In recent years, several studies incorporating mobile devices with printed documents have been carried out. For example, Yeh et al. (2006) developed a mobile learning system which integrated paper notes with in-field ambient information, including digital photos of the surrounding learning targets and the data collected from the real-world environment. Chao and Chen (2009) proposed a learning support system in which a mobile phone was incorporated with printed books and a web-based discussion forum. Bederson, Quinn and Druin (2009) presented a zoomable user interface on mobile phones to support reading scanned multi-lingual printed books. Chen et al. (2011) and Ozcelik

and Acarturk (2011) further demonstrated the use of mobile devices and QR code technology as a complement to printed course materials. A QR Code is a two-dimensional barcode, which consists of black modules arranged in a square pattern on a white background. Within this barcode can be encoded a URL (web address), text, or other information which can be read by a QR code scanner (e.g., Smartphone apps) (Chao & Chen, 2009; Chen, et al., 2011). QR code technology can link offline information to online content, effectively providing additional information and even multimedia resources (Chao & Chen, 2009; Chen, et al., 2011; Ozcelik & Acarturk, 2011).

The previous research not only demonstrated the potential of using mobile technologies to support printed book reading, but also revealed the need to develop adequate Mindtools or knowledge construction tools to assist students in organizing knowledge obtained from books (Liu, Chen, & Chang, 2010). For example, Hwang, Shi and Chu (2011) have reported that the students who learned with concept maps in ubiquitous learning environments have significantly better learning outcomes than those who learn with the traditional tour-based ubiquitous learning approach. The study of Hung, Hwang, Su, and Lin (2012) in an in-field ecology observation activity has also shown that the use of concept maps in ubiquitous learning activities is helpful to students in improving their learning outcomes. Consequently, in this study, we have developed a ubiquitous learning system to support printed science book reading by providing supplementary learning materials to complement the printed content, and a concept map facility to assist students in organizing their knowledge learned from the books.

### **Computerized concept maps**

Novak and Gowin (1984) have indicated that meaningful learning can be formed by helping students connect their old cognitive structures with the new ones during the teaching process. Among various knowledge organizing tools, concept maps have been recognized by researchers as being an effective tool for externalizing students' knowledge and reorganizing both their old and new knowledge (Akinsanya & Williams, 2004; Hwang, Shi, & Chu, 2011). In the past decades, concept maps have been used as a knowledge organizing tool for supporting meaningful learning by helping students represent and visualize their knowledge structures (Hwang, Wu, & Chen, 2012; Lim, Lee, & Grabowski, 2009; Novak, 2002; Trundle & Bell, 2010).

The popularity of computer technology has brought the use of concept maps into the computerized era. Several studies have shown the effectiveness of computerized concept mapping systems in comparison with paper-and-pencil concept maps (Liu, Chen, & Chang, 2010; Kim & Olaciregui, 2008; Wu, Hwang, Milrad, Ke, & Huang, 2012). Researchers have indicated that the computerized approach has simplified the process of creating and revising concept maps; moreover, it has enabled more flexible presentations of the learning content as well as interactions among teachers and students (Hwang, Wu, & Ke, 2011; Reader & Hammond, 1994; Shin et al., 2000). Kim and Olaciregui (2008) have further reported that students seem to recall more ideas when they learn from a concept map than from a textual presentation.

Computerized concept maps have played different roles in various applications in the past decade, including the roles of knowledge construction and organizing tools (Erdogan, 2009), knowledge structure measuring tools (Chang, Sung, Chang, & Lin, 2005), and learning guidance tools (Hwang, 2003; Panjaburee et al., 2010).

### **Concept map-oriented ubiquitous learning system for supporting science book reading (CMULS)**

Figure 1 shows the structure of the concept map-oriented ubiquitous learning system for supporting printed book reading. The ubiquitous learning system consists of three parts: printed science books, mobile devices, and learning resources (i.e., digital learning materials and concept maps). The printed science books are provided with QR codes, which are a type of bar-code containing web addresses of supplementary materials. By using smartphones with wireless communication, learners can scan the QR codes on the printed books to access additional supplementary learning resources from the learning system, including concept maps and learning materials.

Figure 2 shows an example of a printed science book with QR codes. The title of the book is "Lost in the Solar System" (Cole & Degen, 1990), from the Magic School Bus series, which tells the story of a class field trip into outer space during which the students visit each planet in the solar system. By using the smartphone "photo" function to read the QR codes on the book, the corresponding web addresses are accessed and the learning materials as well as the concept maps are thereby presented to the learners.

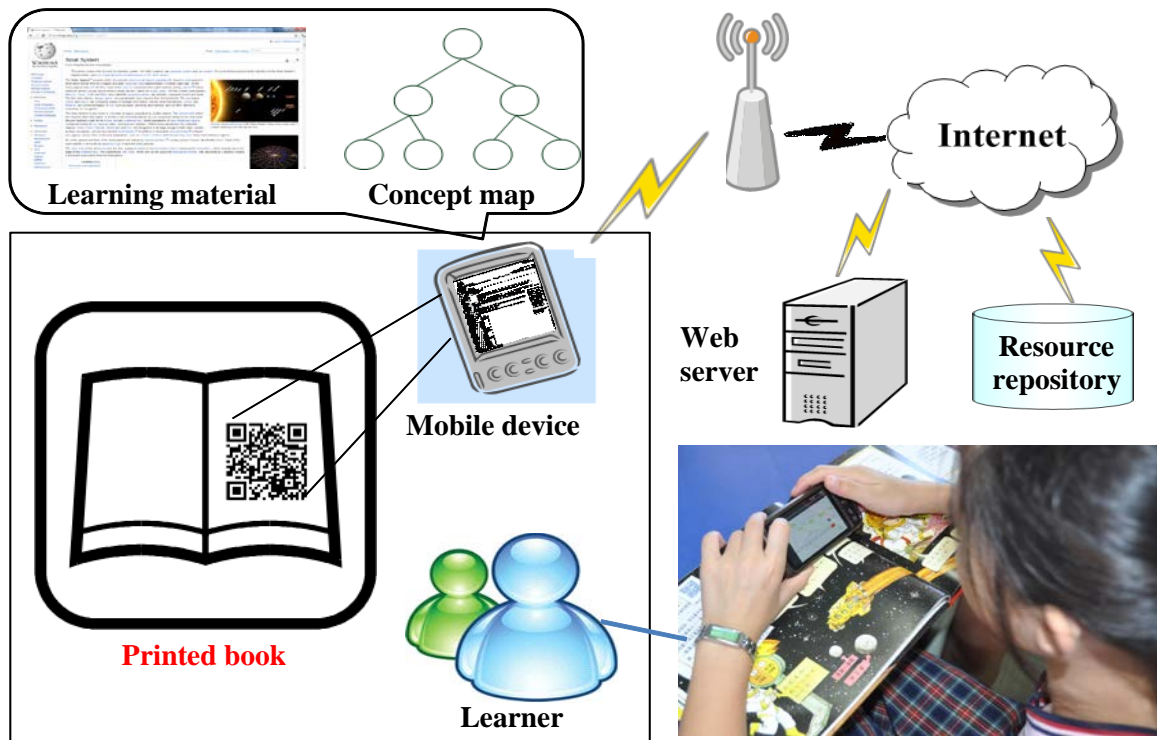


Figure 1. System framework of CMULS



Figure 2. Illustrative example of reading printed science books with concept maps and supplementary materials using mobile devices

In this illustrative example, the concept map is used to help the students organize the information, identify the important concepts and the relationships between concepts, and perceive their learning status related to the "solar system" unit in the science book via visualizing the learning content and learning status information. In the upper right of Figure 2, the concept map shows different planets (in order of distance from the sun). The different colors in the bottom of the screenshot represent the reading progress status. The red node is the planet (e.g. Saturn) that is introduced on the current page, the light blue nodes are the planets that were introduced on the previous pages, and the grey nodes are the planets that will be introduced in the following pages. The left of the screenshot is a QR code icon. By touching the icon, a built-in camera is directly activated and learners can move on to scan other QR codes. The bottom right of Figure 2 shows the learning scenario of reading printed science books with concept maps and supplementary materials using mobile devices.

The supplementary learning materials as well as the demonstrative concept maps were developed by consulting two teachers each with more than five years experience teaching natural science courses and who were informed of the importance of designing term/phenomenon explanations and more realistic illustrations (Ganea, et al., 2008; Tare, et al., 2010). Accordingly, such additional learning materials were developed, including digital text, images, and video clips, to help the students comprehend the learning content. Moreover, different types of learning supports can be accessed via the different colored QR-codes. For example, the blue QR codes link to concept maps (upper right of the figure) while the black codes link to learning materials (bottom right of the figure).

## Research design

This study aims to investigate the effects of the concept map-based u-learning approach on students' learning achievement, cognitive load, and technology acceptance while reading printed science books with handheld devices. The learning content is the subject unit "the Solar System" of an elementary school natural science course.

## Participants

The participants were 92 sixth graders from three classes of an elementary school in southern Taiwan. Two classes were assigned to be Experimental group 1 (N=28) and Experimental group 2 (N=31), while the third class was the control group (N=33). Note that all participants used the same books. The difference between the three groups was the learning treatments. The students in Experimental group 1 were instructed and guided to read printed books with access to concept maps and supplementary learning materials via smart phones; those in Experimental group 2 read printed books with access to supplementary learning materials via smart phones, but without access to concept maps; and those in the control group read the printed books with those supplementary digital materials presented by the teacher. It should be noted that the participants in Experimental groups 1 and 2 did not have a teacher providing further instruction.

## Experimental procedure

Figure 3 shows the experimental procedure, which consists of three stages, that is, conducting the pre-tests, introduction to the tools and learning missions, and conducting the post-tests and the post-questionnaire.

In the first stage, all of the students took the natural science course pre-test. The total time for this stage was 30 minutes. In the second stage, the students in the three groups were instructed with the tools and missions of the learning activity. The time for this stage was 10 minutes.

Following the instruction, a 60-minute learning activity was conducted. During this activity, the students in Experimental group 1 were guided to read the printed books with access to the corresponding concept maps and supplementary learning materials via handheld devices. The students in Experimental group 2 were supplied with supplementary learning materials via handheld devices. On the other hand, the students in the control group used the traditional approach to reading printed books with instruction from the teacher.

In the final stage, the students completed a cognitive load questionnaire, the technology acceptance model

questionnaire, and the natural science course post-test. The total time for this stage was 40 minutes.

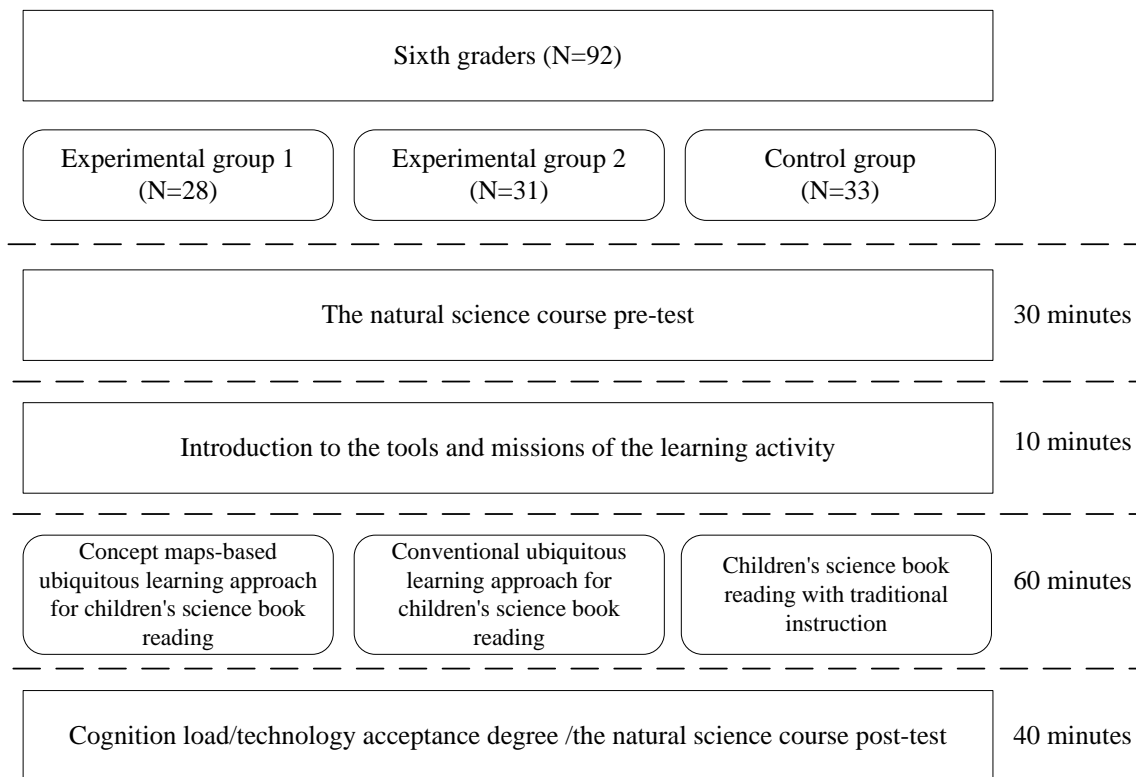


Figure 3. Experimental procedure

### Measuring tools

The pre-test was conducted to evaluate the students' prior knowledge before reading the printed science books. It consisted of 15 true-false questions, 15 multiple-choice questions, 4 matching questions and 5 fill-in-the-blank questions, giving a total score of 100. For example, two of the pre-test items are "Why are we unable to see the constellations of Scorpio and Orion at the same time? (1) Because Scorpio is much brighter than Orion; (2) Because the two constellations move to the same area at different times; (3) Because Scorpio moves faster than Orion; (4) Because Scorpio is closer to the earth" and "When we operate the planisphere (a device to show the locations of constellations), the position of Polaris changes from time to time (1) Yes; (2) No."

The post-test aimed to evaluate the learning achievements of the students after using different treatments to read the printed science books. It consisted of ten multiple-choice questions, with a total score of 10, about the subject unit "the Solar System". The post-test was developed by consulting two teachers who had taught the natural science course for more than five years. For example, two of the post-test items are "Which planet in the Solar System is farthest from the earth? (a) Neptune; (b) Uranus; (c) Mars; (d) Venus; (e) Mercury" and "Why is the temperature on Venus higher than that of the other planets in the Solar System? (a) Because of the ozone layer depletion; (b) Because of the greenhouse effect; (c) Because of its shortest distance to the Sun; (d) Because of the effect of its satellites; (e) Because of its high moving speed."

The cognitive load questionnaire was developed based on the cognitive load measure proposed by Sweller, van Merriënboer and Paas (1998). It consisted of 4 questionnaire items (e.g., "I feel great pressure in learning with this instructional approach" and "I need to put lots of efforts into understanding the learning content") with a 7-point Likert scale, where '7' represented 'strongly agree' and '1' represented 'strongly disagree.' The greater the cognitive burden, the lower the user's satisfaction (Segall, Doolen, & Porter, 2005). The Cronbach's alpha value of the questionnaire was 0.88, showing adequate internal consistency in evaluating the cognitive load of the students.

The questionnaire of technology acceptance was modified from the questionnaire items developed by Davis (1989). It was used to explore how students came to accept and use the mobile technology while reading the printed books with a 7-point Likert scale, where '7' represented 'strongly agree' and '1' represented 'strongly disagree.' The questionnaire included three subscales: 6 items for "usefulness of reading printed books with support from the mobile learning system" (e.g., "The supplementary materials provided by the learning system are quite clear and effectively assist me in understanding the learning content" and "Combining the smartphones and the printed book is helpful to learning"), 7 items for "ease of use" (e.g., "It is easy to operate the learning system via the smartphones and QR codes" and "I learned to operate the learning system quickly") and 6 items for "attitude and intention of future use" (e.g., "I like to learn in this way" and "I would like to recommend this learning system to others"). The Cronbach's alpha values of those three subscales were 0.94, 0.95 and 0.91, respectively, indicating high internal consistency. The students in both of the experimental groups were asked to complete this questionnaire after the learning activity.

## Results

### Learning achievement

Before the experiment, a test of basic science knowledge was conducted to understand the differences in science cognition among the three groups. The descriptive statistics of the pre-test are presented in Table 1.

Table 1. Pre-test results of the natural science

	Group	N	Mean	SD
(E1)	Experimental Group 1	28	83.29	7.12
(E2)	Experimental Group 2	31	92.61	5.18
(C)	Control Group	33	79.24	10.29

After the learning activity, the three groups of students took the post-test. The pre-test scores were regarded as the covariance for analysis of covariance to delete the effects of the pretest on the learning outcome. The homogeneity of the regression coefficient was tested, which revealed that interaction  $F$  between the covariance was 0.31 ( $p > 0.05$ ). This confirms the hypothesis of homogeneity of the regression coefficient.

Table 2 shows the ANCOVA result on the post-test scores of the three groups. The means and standard deviations of the post-test scores were 7.23 and 1.05 for Experimental group 1, 6.61 and 1.50 for Experimental group 2, and 5.73 and 1.64 for the control group. It is found that the post-test scores of the three groups are significantly different, with  $F = 7.80$  ( $p < .01$ ). Moreover, the adjusted means of the three groups were 7.34, 6.37, and 5.91. The pairwise comparisons show that there is a significant difference between E1 and E2, and a significant difference between E1 and C. In other words, the students in Experimental group 1 had significantly better learning achievement than the students in both Experimental group 2 and in the control group.

Consequently, it is concluded that a concept map-based ubiquitous learning approach to reading printed science books with handheld devices is more helpful to the students in terms of learning achievement in a natural science course than traditional printed science book reading and the conventional mobile technology-supported approach, implying the benefit of leading in with the concept maps to assist students to learn via mobile devices.

Table2. ANCOVA result of learning achievement on the post-test scores of the three groups

Group	N	Mean	S.D.	Adjusted Mean	F(2, 88)	Pairwise comparisons
(E1) Experimental Group 1	28	7.23	1.05	7.34	7.80**	
(E2) Experimental Group 2	31	6.61	1.50	6.37		(E1)>(E2)
(C) Control Group	33	5.73	1.64	5.91		(E1)>(C)
Total number of students	92	6.50	1.56			

\*\* $p < .01$

## Cognitive load

Table 3 presents the analysis result of the students' cognitive load. The means and standard deviations are 3.55 and 1.36 for the control group, 3.86 and 1.38 for Experimental group 1, and 4.35 and 1.94 for Experimental group 2. Although both the experimental groups revealed slightly higher mean scores than the control group, the ANOVA result ( $F = 2.08$  and  $p > .05$ ) showed no significant difference between the three groups, implying that the three groups of students experienced equivalent levels of cognitive load during the learning activity. Moreover, the average cognitive loads of the three groups were not high, implying that printed science books provide an easy and relaxed way for students to read. In addition, it is interesting to find that Experimental group 1 had lower average cognitive load than Experimental group 2, showing that the lead in of the concept maps could have eased the cognitive load of the students in terms of using the mobile devices to access the supplementary digital materials.

*Table 3. ANOVA result of cognitive load on the three groups*

	Group	N	Mean	S.D.	F(2, 89)
(E1)	Experimental Group 1	28	3.86	1.38	2.08
(E2)	Experimental Group 2	31	4.35	1.94	
(C)	Control Group	33	3.55	1.36	

## Acceptance of using the mobile technology to support printed science book reading

Table 4 presents an analysis of the students' degree of acceptance of using the mobile technology to support printed science book reading. The means and standard deviations of the usefulness subscale are 5.95 and 0.97 for Experimental group 1 and 6.24 and 1.12 for Experimental group 2, showing that most students felt that the proposed system was useful. From the interview, it was found that the students in both experimental groups showed positive attitudes toward reading the supplementary materials on the smartphones. Some of the students in Experimental group 1 further revealed that they were highly appreciative of the concept maps helping them understand the relationships between the learning items in the books.

In terms of ease of use, the means and standard deviations are 5.48 and 1.18 for Experimental group 1, and 6.04 and 1.18 for Experimental group 2; that is, most students felt that the ubiquitous learning system was easy to use. Moreover, the scores of the two groups were not significantly different, implying that the students in Experimental group 1 did not feel that it was difficult to receive the learning supports although they needed to read the concept maps via the handheld devices.

In terms of "attitude and intention of future use," the means and standard deviations are 5.90 and 0.88 for Experimental group 1, and 6.19 and 0.99 for Experimental group 2. It was found that the  $t$ -test result shows no significant difference between the two groups, implying that the students in both experimental groups revealed similar degrees of acceptance of the ubiquitous learning system. Furthermore, from the interview, it is interesting to find that most of the students felt that the use of the QR-codes was convenient, efficient and interesting, although there were alternative ways of accessing the digital materials, such as selecting the items on the menu or searching for the data with keywords. This implies that the provision of a user-friendly interface is able to promote the willingness of students to use the learning system.

*Table 4.  $t$ -test result of technology acceptance of the experimental groups*

Scale		Group	N	Mean	S.D.	$t$
Usefulness	(E1)	Experimental Group 1	28	5.95	0.97	1.06
	(E2)	Experimental Group 2	31	6.24	1.12	
Ease of Use	(E1)	Experimental Group 1	28	5.48	1.18	1.82
	(E2)	Experimental Group 2	31	6.04	1.18	
Attitude	(E1)	Experimental Group 1	28	5.90	0.88	1.18
	(E2)	Experimental Group 2	31	6.19	0.99	



## Discussion

To achieve the aim of assisting children in printed science book reading, a concept map-based ubiquitous learning approach was proposed in this study. The experimental results showed that the proposed approach had significantly better effectiveness in improving students' learning achievements than the conventional ubiquitous learning approach and the traditional printed science book reading approach. In addition, the experimental results showed no significant difference between Experimental group 2 and the Control group; that is, the effect of accessing digital supplementary materials was equivalent to that of the teacher's instruction. Such a finding conforms to what has been reported by previous studies (Chen et al., 2011). Meanwhile, the analysis of the questionnaire results showed that the proposed learning approach did not increase the students' cognitive burden; moreover, it can be seen from Table 4 that most of the students held positive views regarding "ease of use," "usefulness" and "attitude and intention of future use" regarding the ubiquitous reading-supported system.

To sum up, the proposed approach is helpful to students in terms of learning achievement, cognitive load and technology acceptance. Consequently, to effectively support printed science book reading, the provision of both concept maps and digital supplementary materials is required. Such an approach can also be applied to the printed book reading of other subjects, such as mathematics, language, social science, and natural science, in which concept maps could be helpful to readers in organizing the concepts in the book content, and hence their reading comprehension can be improved.

## Conclusions

This study investigates the effects of a concept map-based ubiquitous learning approach to reading printed science books with handheld devices on students' learning achievement, cognitive load, and technology acceptance. A learning activity has been conducted to compare the learning performance of students who read printed science books with mobile device-supported concept maps and supplemental learning materials, those who read printed science books with access to digital supplemental learning materials only, and those who read printed science books with traditional instruction. A quasi-experiment has been conducted in an elementary school to evaluate the effectiveness of the proposed approach. The experimental results show that the concept map-oriented ubiquitous learning approach is significantly more helpful to the students in reading printed books than traditional book reading and the conventional ubiquitous learning approach in terms of learning achievements; moreover, the students had a high level of acceptance of such a mobile technology-assisted learning system in terms of "ease of use", "usefulness" and "attitude and intention of future use."

In sum, the experimental results show the importance of providing learning supports for printed science book reading. Therefore, one of the salient contributions of this study is to highlight the importance of providing concept maps and supplementary materials for printed science book reading. Another contribution is to demonstrate how the proposed approach can be used in a specific learning activity and to show its potential in terms of its effectiveness in improving the learning achievements and technology-acceptance degrees of the students.

Although the experimental results have shown the benefits of using CMULS, there are some limitations in the present study. For example, although mobile devices such as smartphones are becoming increasingly popular, not all students have access to such a learning device at present. Fortunately, researchers have predicted that mobile devices will become a common learning device in the near future (Norris, Hossain, & Soloway, 2011), implying that the proposed approach has the potential to become a widely-used learning model. Another limitation of this study is that the QR codes on the book are fixed. Even though concept maps as well as supplementary learning materials are offered, students may still experience difficulty reading the book. That is, adaptive learning materials for individual students to complement the printed content are needed. Therefore, one of our future works is to extend this study to offer adaptive learning resources for students.

## Acknowledgements

This study is supported in part by the National Science Council of the Republic of China under contract numbers NSC 99-2511-S-011-011-MY3 and NSC 101-2511-S-011 -005 -MY3.

## References

- Ackerman, R., & Goldsmith, M. (2008). Learning directly from screen? Oh-no, I must print it! Metacognitive analysis of digitally presented text learning. In Y. Eshet, A. Caspi, & N. Geri (Eds.), *Proceedings of the Chais Conference on Instructional Technologies Research 2008: Learning in the Technological Era* (pp. 1-7). Raanana, Israel: The Open University of Israel.
- Akinsanya, C., & Williams, M. (2004). Concept mapping for meaningful learning. *Nurse Education Today*, 24(1), 41-46.
- Allor, J. H., & McCathren, R. B. (2003). Developing emergent literacy skills through storybook reading. *Intervention in School and Clinic*, 39(2), 72-79.
- Bederson, B. B., Quinn, A., & Druin, A. (2009). Designing the reading experience for scanned multi-lingual picture books on mobile phones. In Heath, F., Rice-Lively, M. L., & Furuta, R. (Eds.), *Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries* (pp. 305-308): New York, NY: ACM Press.
- Chan, T., Roschelle, J., Hsi, S., Kinshuk, Sharples, M., Brown, T., ... Hoppe, U. (2006). One-to-one technology-enhanced learning: An opportunity for global research collaboration. *Research and Practice in Technology Enhanced Learning*, 1(1), 3-29.
- Chang, K. E., Lan, Y. J., Chang, C. M., & Sung, Y. T. (2010). Mobile device-supported strategy for Chinese reading comprehension. *Innovations in Education and Teaching International*, 47(1), 69-84.
- Chang, K. E., Sung, Y. T., Chang, R. B., & Lin, S. C. (2005). A new assessment for computer-based concept mapping. *Educational Technology & Society*, 8(3), 138-148.
- Chao, P. Y., & Chen, G. D. (2009). Augmenting paper-based learning with mobile phones. *Interacting with Computers*, 21(3), 173-185.
- Chen, N.-S., Teng, D. C.-E., Lee, C.-H., & Kinshuk. (2011). Augmenting paper-based reading activity with direct access to digital materials and scaffolded questioning. *Computers & Education*, 57(2), 1705-1715.
- Chiou, C. K., Tseng, J. C. R., Hwang, G. J., & Heller, S. (2010). An adaptive navigation support system for conducting context-aware ubiquitous learning in museums. *Computers & Education*, 55(2), 834-845.
- Chu, H. C., Hwang, G. J., & Tsai, C. C. (2010). A knowledge engineering approach to developing mindtools for context-aware ubiquitous learning. *Computers & Education*, 54(1), 289-297.
- Chu, H. C., Hwang, G. J., & Tseng, J. C. R. (2010). An innovative approach for developing and employing electronic libraries to support context-aware ubiquitous learning. *Electronic Library*, 28(6), 873-890.
- Cole, J., & Degen, B. (1990). *The magic school bus, lost in the solar system*. New York, NY: Scholastic Inc.
- Coulby, C., Hennessey, S., Davies, N., & Fuller, R. (2011). The use of mobile technology for work-based assessment: The student experience. *British Journal of Educational Technology*, 42(2), 251-265.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *Mis Quarterly*, 13(3), 319-340.
- Dünser, A., & Hornecker, E. (2007). An observational study of children interacting with an augmented story book. In K. Hui et al. (Eds.), *Proceedings of the 2nd International Conference on Technologies for E-learning and Digital Entertainment* (pp. 305-315). doi: 10.1007/978-3-540-73011-8\_31
- Erdogan, Y. (2009). Paper-based and computer-based concept mappings: The effects on computer achievement, computer anxiety and computer attitude. *British Journal of Educational Technology*, 40, 821-836.
- Fletcher, K. L., & Reese, E. (2005). Picture book reading with young children: A conceptual framework. *Developmental Review*, 25(1), 64-103.
- Ganea, P. A., Pickard, M. B., & DeLoache, J. S. (2008). Transfer between picture books and the real world by very young children. *Journal of Cognition and Development*, 9(1), 46-66.
- Garza, T. J. (1991). Evaluating the use of captioned video materials in advanced foreign language learning. *Foreign Language Annals*, 24(3), 239-258.
- Hoven, D. (1999). A model for listening and viewing comprehension in multimedia environments. *Language Learning & Technology*, 3 (1), 88-103.
- Hung, P. H., Hwang, G. J., Su, I. H., & Lin, I. H. (2012). A concept-map integrated dynamic assessment system for improving ecology observation competences in mobile learning activities. *Turkish Online Journal of Educational Technology*, 11(1), 10-19.
- Hwang, G. J. (2003). A concept map model for developing intelligent tutoring systems. *Computers & Education*, 40(3), 217 - 235.
- Hwang, G. J., & Chang, H. F. (2011). A formative assessment-based mobile learning approach to improving the learning attitudes

and achievements of students. *Computers & Education*, 56(4), 1023-1031.

Hwang, G. J., Chu, H. C., Lin, Y. S., & Tsai, C. C. (2011). A knowledge acquisition approach to developing Mindtools for organizing and sharing differentiating knowledge in a ubiquitous learning environment. *Computers & Education*, 57(1), 1368-1377.

Hwang, G. J., Chu, H. C., Shih, J. L., Huang, S. H., & Tsai, C. C. (2010). A decision-tree-oriented guidance mechanism for conducting nature science observation activities in a context-aware ubiquitous learning environment. *Educational Technology & Society*, 13(2), 53-64.

Hwang, G. J., Shi, Y. R., & Chu, H. C. (2011). A concept map approach to developing collaborative Mindtools for context-aware ubiquitous learning. *British Journal of Educational Technology*, 42(5), 778-789.

Hwang, G. J., Wu, P. H., & Ke, H. R. (2011). An interactive concept map approach to supporting mobile learning activities for natural science courses. *Computers & Education*, 57(4), 2272-2280.

Hwang, G. J., Tsai, C. C., Chu, H. C., Kinshuk, & Chen, C. Y. (2012). A context-aware ubiquitous learning approach to conducting scientific inquiry activities in a science park. *Australasian Journal of Educational Technology*, 28(5), 931-947.

Hwang, G. J., Tsai, C. C., & Yang, S. J. H. (2008). Criteria, strategies and research issues of context-aware ubiquitous learning. *Educational Technology & Society*, 11(2), 81-91.

Hwang, G. J., Wu, P. H., & Chen, C. C. (2012). An online game approach for improving students' learning performance in web-based problem-solving activities. *Computers & Education*, 59(4), 1246-1256.

Kim, P., & Olaciregui, C. (2008). The effects of a concept map-based information display in an electronic portfolio system on information processing and retention in a fifth-grade science class covering the Earth's atmosphere. *British Journal of Educational Technology*, 39, 700-714.

Kuo, F. R., Hwang, G. J., & Lee, C. C. (2012). A hybrid approach to promoting students' web-based problem solving competence and learning attitude. *Computers & Education*, 58(1), 351-364.

Lim, K. Y., Lee, H. W., & Grabowski, B. (2009). Does concept-mapping strategy work for everyone? The levels of generativity and learners' self-regulated learning skills. *British Journal of Educational Technology*, 40(4), 606-618.

Liu, P. L., Chen, C. J., & Chang, Y. J. (2010). Effects of a computer-assisted concept mapping learning strategy on EFL college students' English reading comprehension. *Computers & Education*, 54(2), 436-445.

Liu, T. C., Peng, H. Y., Wu, W. H., & Lin, M. S. (2009). The effects of mobile natural-science learning based on the 5E learning cycle: A Case Study. *Educational Technology & Society*, 12(4), 344-358.

Looi, C.-K., Seow, P., Zhang, B., So, H.-J., Chen, W., & Wong, L.-H. (2010). Leveraging mobile technology for sustainable seamless learning: A research agenda. *British Journal of Educational Technology*, 41(2), 154-169.

Looi, C.-K., Wong, L.-H., So, H.-J., Seow, P., Toh, Y., Chen, W., ... Soloway, E. (2009) Anatomy of a mobilized lesson: Learning my way. *Computers and Education*, 53, 1120-1132.

Mantzicopoulos, P., & Patrick, H. (2010). "The Seesaw is a machine that goes up and down": Young children's narrative responses to science-related informational Text. *Early Education and Development*, 21(3), 412-444.

Mantzicopoulos, P., & Patrick, H. (2011). Reading picture books and learning science: Engaging young children with informational text. *Theory into Practice*, 50(4), 269-276.

McKenzie, J., & Darnell, D. (2003). *The eyeMagic book. A report into augmented reality storytelling in the context of a children's workshop*. Retrieved from <http://www.mindspacesolutions.com/demos/eyeMagicWorkShopReport.pdf>

Norris, C., Hossain, A., & Soloway, E. (2011). Using smartphones as essential tools for learning- A call to place schools on the right side of the 21st century. *Educational Technology*, 18-25. Retrieved from <http://cecs5580.pbworks.com/w/file/50303907/ET%20article%20Norris%20Soloway.pdf>

Novak, J. D. (2002). Meaningful learning: The essential factor for conceptual change in limited or appropriate propositional hierarchies (LIPHs) leading to empowerment of learners. *Science Education*, 86(4), 548-571.

Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. Cambridge, NY: Cambridge University Press.

Ogata, H., Matsuka, Y., El-Bishouty, M. M., & Yano, Y. (2009). LORAMS: linking physical objects and videos for capturing and sharing learning experiences towards ubiquitous learning. *International Journal of Mobile Learning and Organisation*, 3(4), 337-350.

O'Hara, K., & Sellen, A. (1997). A comparison of reading paper and online documents. In O'Hara, K., & Sellen, A. (Eds.), *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '97)* (pp. 335-342). New York, NY: ACM Press.

- Ortiz, C., Stowe, R. M., & Arnold, D. H. (2001). Parental influence on child interest in shared picture book reading. *Early Childhood Research Quarterly*, 16(2), 263-281.
- Ozcelik, E., & Acarturk, C. (2011). Reducing the spatial distance between printed and online information sources by means of mobile technology enhances learning: Using 2D barcodes. *Computers & Education*, 57(3), 2077-2085.
- Panjaburee, P., Hwang, G. J., Triampo, W., & Shih, B. Y. (2010). A multi-expert approach for developing testing and diagnostic systems based on the concept effect model. *Computers & Education*, 55, 527-540.
- Paris, A. H., & Paris, S. G. (2003). Assessing narrative comprehension in young children. *Reading Research Quarterly*, 38(1), 36-76.
- Peng, H. Y., Chuang, P. Y., Hwang, G. J., Chu, H. C., Wu, T. T., & Huang, S. X. (2009). Ubiquitous performance-support system as Mindtool: A case study of instructional decision making and learning assistant. *Educational Technology & Society*, 12(1), 107-120.
- Reader, W., & Hammond, N. (1994). Computer-based tools to support learning from hypertext: concept mapping tools and beyond. *Computers & Education*, 12, 99-106.
- Rogers, Y., Price, S., Randell, C., Fraser, D. S., Weal, M., & Fitzpatrick, G. (2005). Ubi-learning integrates indoor and outdoor experiences. *Communications of the ACM*, 48(1), 55-59.
- Roschelle, J., & Pea, R. (2002). A walk on the WILD side: How wireless handhelds may change computer-supported collaborative learning. *International Journal of Cognition and Technology*, 1(1), 145-168.
- Segall, N., Doolen, T. L., & Porter, J. D. (2005). A usability comparison of PDA-based quizzes and paper-and-pencil quizzes. *Computers & Education*, 45(4), 417-432.
- Shih, J. L., Chuang, C. W., & Hwang, G. J. (2010). An inquiry-based mobile learning approach to enhancing social science learning effectiveness. *Educational Technology & Society*, 13(4), 50-62.
- Shin, J., Deno, S. L., Robinson, S. L., & Marston, D. (2000). Predicting classroom achievement from active responding on a computer-based groupware system. *Remedial and Special Education*, 21(1), 53-60.
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251-296.
- Tare, M., Chiong, C., Ganea, P., & DeLoache, J. (2010). Less is more: How manipulative features affect children's learning from picture books. *Journal of Applied Developmental Psychology*, 31(5), 395-400.
- Trundle, K. C., & Bell, R. L. (2010). The use of a computer simulation to promote conceptual change: A quasi-experimental study. *Computers & Education*, 54(4), 1078-1088.
- Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development*, 69(3), 848-872.
- Whitehurst, G. J., Falco, F. L., Lonigan, C. J., Fischel, J. E., Debaryshe, B. D., Valdezmenchaca, M. C., & Caulfield, M. (1988). Accelerating language-development through picture book reading. *Developmental Psychology*, 24(4), 552-559.
- Wong, L. H., Chin, C. K., Tan, C. L., & Liu, M. (2010). Students' personal and social meaning making in a Chinese idiom mobile learning environment. *Educational Technology & Society*, 13(4), 15-26.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of child psychology and psychiatry, and allied disciplines*, 17(2), 89-100.
- Wu, P. H., Hwang, G. J., Milrad, M., Ke, H. R., & Huang, Y. M. (2012). An innovative concept map approach for improving students' learning performance with an instant feedback mechanism. *British Journal of Educational Technology*, 43(2), 217-232.
- Yang, J. C., Huang, Y. T., Tsai, C. C., Chung, C. I., & Wu, Y. C. (2009). An automatic multimedia content summarization system for video recommendation. *Educational Technology & Society*, 12 (1), 49-61.
- Yang, S. J. H. (2006). Context aware ubiquitous learning environments for peer-to-peer collaborative learning. *Educational Technology & Society*, 9(1), 188-201.
- Yeh, R. B., Liao, C., Klemmer, S. R., Guimbretière, F., Lee, B., Kakaradov, B., & Paepcke, A. (2006). ButterflyNet: A mobile capture and access system for field biology research. In R. Grinter et al. (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 571-580). New York, NY: ACM.

## On the Educational Validity of Research in Educational Technology

Chris Tompsett

Faculty of Health and Social Care Sciences, Kingston University and St. Georges University of London, Kingston Hill, Kingston-upon-Thames, KT2 7LB, UK // [cptompsett@kingston.ac.uk](mailto:cptompsett@kingston.ac.uk)

(Submitted May 17, 2012; Revised August 07, 2012; Accepted September 24, 2012)

### ABSTRACT

Alsop and Tompsett argued in 2007 that research in educational technology was over-focused on technological design and that empirical research that was pertinent to educational change should be assessed against two distinct scales. The first considers the quality of the experimental design in validating the conclusions that are reached, whilst the second assesses the extent to which the experiment reflects normal practice. They suggested that research should only claim to provide evidence for specific change in educational practice if it met the highest levels on both scales. This study, structured in the form of a systemic review covering more than 40 years of published research, provides the first systematic assessment of their model. This analysis suggests that, in at least one key domain in higher education, there is no current reliable research evidence, published in English, that supports the case for a change in practice based on the introduction of information and communication technology (ICT). In this context, practitioners would appear to be fully justified in considering innovations that do not depend on ICT. If practitioners are expected to change, then research must be designed so that the results are transferable, irrespective of the methodology used.

### Keywords

Systematic review, Level of evidence, Empirical design, Bias, Computer science, Higher education

### Introduction

Despite the advances that have occurred over the last 40 years in the use of technology in education, the impact of this research on educational practice remains limited. Low-technology tools have been replaced with equivalent computer hardware and software, but these changes are common to all organizations, educational or otherwise. More specific proposals for increasing the use of educational technology remain largely unheeded by most of those who teach ‘at the coalface’. Advocates for increased use of technology prefer to attribute the lack of impact to characteristic failings among the practitioners, categorising them as Luddites, technophobes, and/or ‘laggards’ — after Rogers (1962). Yet the persistence of the problem merits deeper consideration. Luddites only exist when the technology is more efficient, and technophobia cannot account for rejection in fields, such as computer science, where knowledge and practice in information and communication technology (ICT) is central to the domain. The classification as ‘laggards’ is also invalid. Roger defines laggards as those who adopt a change long after the majority have done so; the laggards can never be the majority in such circumstances.

Alsop and Tompsett (2007), in their analogy with research and innovation in healthcare from a ‘soft-systems’ (Checkland, 1999) perspective, argued that research in this field is over-focused on technological change, with a corresponding disregard for demonstrating that innovative technology provides assured educational benefits for those who adopt it. If there is insufficient evidence for practitioners to adopt a technological change in practice in their particular domain, then it is rational to consider, and prefer, non-technological changes that could provide more assured benefits and/or introduce fewer risks. The focus on practitioners draws on a key distinction made by Oancea (2005) in the debate over the quality, value, and even feasibility of research in the complex domain of education (see, for example: Hargreaves, 1996; Hammersley, 2002; OECD CERI, 2002; Simons, 2003). Oancea categorises researchers as either ‘intellectuals’, who focus on elucidating educational problems, or ‘technicians’, who focus on providing solutions to known educational problems. Alsop and Tompsett suggested that if research by ‘technicians’ were to be classified as in Table 1, below, then almost all valid research in ICT would be classified at level A. They argued that practitioners (regardless of their own experience and abilities) would require evidence that could be classified at level C or higher to warrant a change in practice.

Taking computer science as a focus, this paper provides the first review of published research focused in one specific domain in higher education. The search for, and selection of relevant evidence mirrors the principles of a systematic review (as described, for example, in Jadad, 1998).

Table 1. Classification of research in ICT in education

Phase	Underlying Research Question
A: Effect	Can a new ICTE part of a course be shown to (1) have an effect (2) have an effect on learning, within a limited number of students in advantageous conditions?
B: Efficacy	Can a new ICTE part of a course be shown to have a positive effect on learning across a suitably large, selected range of students who study properly?
C: Effectiveness	Can a new ICTE part of a course be shown to have a positive effect on learning across a suitably large range of students where no control is maintained on how it is used??
D: Efficiency	Does the introduction of a new effective ICTE part of a course with a limited set of resources, for a specific group of students represent the best use of resources?
E: Side-effects	What otherwise unknown side effects result from full-scale use of a new ICTE component in a course?

Extracted from (Alsop & Tompsett 2007, table 6)

Systematic reviews differ in two critical ways from conventional literature reviews (e.g., Sheard et al. 2009), and/or meta-analysis of existing results (e.g., Lou, Abrami and d'Apollonia, 1996, or Springer, Stanne and Donovan 1999). Firstly, the focus of research must be specific; evidence for change with less able students in mathematics in secondary school cannot be used as evidence for change when teaching at different levels of education or with different levels of ability. In this case, the focus is the teaching of a 'first' computing language in an undergraduate computer-science program. Secondly, the value of an experimental study is assessed primarily in terms of the first scale proposed by Alsop and Tompsett: the quality of evidence that is achieved by the experiment, and comparing those changes that are supported at the highest level of research that can be achieved, ignoring both the specific proposal and the actual improvement in outcome at this point. More specifically, a systematic review follows a five stage process consisting of: (1) collecting evidence for all relevant interventions, (2) evaluating the quality of the research process for each study, (3) elimination of findings that are below the highest achieved within that set of studies, (4) collating the results of homogeneous experiments to increase, if possible, the reliability of evidence for each intervention, and, finally, (5) consideration of the scale and spread of the impact that can be attributed to each remaining intervention in order to identify the 'best' option, or combination of options.

The following section defines the focus for the study and identifies critical variables that are associated with students' learning in this field at this level. These variables will need to be accounted for and controlled within the research that is reviewed. The following two sections follow the five-stage process (outlined above), as far as possible. The final section reviews the outcomes of this analysis. Whilst the approach is based on the principles of a systematic review, as detailed in Higgins and Green (2009), for example, it is not possible to adhere to all of these, and these limits are considered in the final section.

## Researching the problem

This review focuses on the difficulties of teaching a first computing language in higher education, irrespective of the particular model of learning assumed by the original researcher (see, *inter alia*, Mayer, 1975; Awad, 1989; Hagan & Macdonald 2000, and Golding, Donaldson, & Tennant 2009). Failure rates have been noted as high throughout the literature, and there is little evidence of improvement (see: Hagan & Macdonald, 2009; Truong, Bancroft & Roe, 2003; Bornat, Dehnadi & Simon 2008; Smith & Fidge, 2008, and Corney, Teague & Thomas 2010). Even if students succeed in a first course, the experience can act as a deterrent to further study in the field. Any change in practice must demonstrate measurable and sustained improvement in achievement in one or more of the following: acquiring a precise understanding of a specific computer language (syntactic knowledge), developing an accurate mental model of what the computer does (conceptual knowledge), and finally, developing the ability to reflect a 'real world' problem within this mental model, so that the real world problems can be solved within the constraints of the mental model (after Bayman & Mayer, 1988).

A large number of studies have investigated the nature of the problem, rather than its solution. These explore the relationship between overall student achievement and/or progress and indicators such as student aptitude and/or cognitive characteristics. Campbell and McCabe (1984), for example, studied overall retention rates for first year computer science students in their university in the USA. From the 11 variables considered, factors such as gender, and ability in science, mathematics and verbal reasoning, were all highly significant, although the most accurate

model (of 2047 generated) employed only one factor, the average grade in high school mathematics and only predicted those who switched out of computer science as a major. Johnson and Johnson (1992) considered the association between four well-established cognitive characteristics (introversion-extroversion, stability-neuroticism, personal stress, and creativity) and four outcome variables: productivity, programming, conceptual understanding and final grade, identified. They established, that both age and fluency (part of creativity) were significant to predicting programming and final grade (at a significance level below .001), and that gender correlated with productivity, programming and final grade (below .01). The models developed by Nowaczyk (1984); Fletcher (1984); de Raadt et al. (2005); Golding et al. (2009); Ford and Venema (2010) are also based on established measures, but there is only limited consistency between these models (de Raadt, et al., 2005 and Katz, Aronis, Allbritton, et al. 2003). The quality of these papers is variable, however, and too many studies introduce (or redefine) variables without considering the validity of doing so. Pintrich, Smith, Garcia and McKeachie, 1993; Bayman, 1983 and Bayman & Mayer, 1988 would be notable exceptions.

Two issues are pertinent to the analysis of interventions that follows. The first is that there is a core set of variables that should be controlled, or accounted for, within any study. Even though no single model has been developed, the omission of key variables prevents the elimination of bias when comparing the performance of any two groups. The second is that contextual factors that remain constant across a study, such as the profile of the students in the study (as a sub-population of students of the same subject across higher education), limits generalization beyond the immediate study. Indeed, institutional variables may be more significant than those that have been shown to be relevant in a single institution (de Raadt et al., 2005).

### **Possible changes in educational practice in teaching computer science**

Three general categories of intervention have been considered in the papers that were reviewed: a change in the programming language that is taught, changes to course organisation, and changes to the way in which students study. Any one of these may also involve a change in the way students are assessed — either formatively or summatively. Any intervention considered included the application of ICT as a critical factor in the design of the intervention.

The choice of programming language is closely related to the development of appropriate mental models. Apart from the choice between procedural and object-oriented languages (Sheard, Carbone & Markham, 2000), a range of simplified models have been proposed, such as the use of mini-languages with more accessible mental models (e.g., Cooper, Dann, and Pausch, 2000; Carlisle, Wilson, Humphries, and Hadfield 2005; Crawford & Boese, 2006), languages for gaming devices (Tew, Dorn, Leahy, and Guzdial 2008), and cut-down versions of full languages (Brusilovsky et al., 1997). In each case, however, it should be evident that benefits can be transferred to the main programming language used later in a course.

Changes in organisation focus on the way in which the resources, such as staff and students are provided, while excluding those that simply increase non-technical resources, or that change the underlying student population. Some changes can be endemic. The programming environment affects a wide range of factors that can support the student in developing competence in constructing/correcting a computer program, as well as provide support systems that enable, for example, the submission and automated marking of coursework (Reid & Wilson, 2005), or provide more formative feedback (Johnson & Soloway, 1985). The introduction of an intelligent tutoring application would fall within this category.

Changes to patterns of study covers issues such as required attendance at workshops (Thweatt, 1994), the introduction of paired/team working (e.g., Cheney, 1977; Williams et al. 2002; and Serbec, Kaucic & Rugelj 2008), but they would only be included if improvements could be attributed to the use of ICT, rather than to the underlying change in educational delivery.

Changes can, of course, affect several aspects at the same time. Changing the language and programming environment will often require changes in course structure—see the study by Bloch (2000), and the use of a mini-language (Carlisle et al., 2005) required an integrated set of changes in all aspects.

## **Constraints**

A full systematic review formalizes the assessment of evidence as an objective process conducted by a team of experts. This implies that each stage of the process is replicable, to a given level of accuracy, by other researchers. The resources to conduct this research with a team have not been available. Nevertheless, the resources that were used have exceeded, by far, those that a practitioner could afford in considering a change in their practice.

Three constraints limited this search process. Firstly, the search for possible evidence was constrained to those resources that were identifiable and published before July 2011. An earliest publication date was taken as the beginning of January 1966, as the majority of studies before that date were limited to the programmed learning method, which is now seem as over-limited even by those in this field (Emurian, Hu, Yang & Durham 2000). Secondly, the relevant source needed to be accessible, in English, through the UK library system by the end of April 2011. The most apparent effect is an inability to access material from early conferences and PhD dissertations that are not available in electronic form. The effect of restricting the search to English language publications cannot be assessed.

## **Summary**

The search process is designed to identify any individual studies that report the outcomes of a change in practice in teaching a first computer language, based on the introduction of ICT, at the undergraduate degree level, to students who need to develop sufficient understanding of computer programming to continue studies in the field. Individual studies are identified by the set of subjects on which the report is based. The following section provides a more detailed description of the search process.

## **Collecting evidence**

The search for relevant studies was formalized as two processes, an inclusion and exclusion phase, conducted in parallel. The inclusion phase is used to identify all relevant studies from papers by setting weak criteria for selection of papers; the exclusion phase is conducted by excluding any of these studies unless it meets all of a number of strictly-defined criteria.

### **Inclusion phase**

The papers for inclusion were selected in three stages. The first stage took data from two meta-studies by Springer et al. (1999) and Lou, Abrami & d'Apollonia (1996) in order to pilot that the process for tracing sources and recording the relevant information. All records entered in this stage were subsequently reassessed against the standard inclusion/ exclusion criteria to ensure consistency.

The second stage was conducted using standard abstracting systems for both higher education and computing (ERIC, AUEA, BEI, ACM-Guide, ERA, and UMI Dissertations Index). The index terms varied, as they reflected the indices available. The third stage identified key papers that referenced, or were referenced by, papers that had already been included.

The initial search was conducted in September 2008 and updated in July 2011. More than 800 references were identified and accessed by the end of this phase.

### **Exclusion phase**

The exclusion phase was conducted in two passes. In the first pass, the abstracts, and then the full paper for each reference was reviewed in order to identify all relevant studies, irrespective (more or less) of the research model that was applied. In the second pass, each study was reviewed in detail in order to ensure that they were relevant to the specific educational focus of the study.



## First pass

Articles were excluded unless they met the following four conditions:

- The article should report on at least one study related to a change in practice and include experimental data.
- An underlying population (of students) should be defined that included an identifiable subset of subjects relevant to computer science, or a related subject (software engineering, information systems), requiring at least one follow-on course in computer programming.
- The process for identifying each group and subgroup of students should be clearly defined.
- The underlying research model should be generalizable.

The final condition excluded only two small sets of studies: those where the underlying model of research, such as action research, is inherently context-dependent (see, for example, Falkner & Palmer, 2009), and two studies where the authors state that the results are non-generalizable—Ophel, Prosser, and Robinson 1997, and Sajaniemi, Kuittinen, and Tikansola, 2008. Exclusion under this condition implies only that there is no basis for assuming that differential outcomes could be transferred to a different environmental context — see Lewis, Chase, Pérez-Quñones, and Rosson (2005). Studies that introduced equivalent technology for all groups (e.g., Thornburg, 2010), or that introduced technology for recording and research purposes alone (e.g., Boyer et al. 2010) also were excluded.

Four references could not be accessed as papers. Burnett (1979) was based on teaching FORTRAN, and this is seldom used as a first computer language except in engineering. The remaining three (Chen & Lin 2006; Li, Wang & Zhang, 2006, and Tan, Ting & Ling, 2009) do not appear to include experimental data.

## Second pass

The following six criteria were used to ensure that each study was directly relevant to the problem being considered:

- The subjects should be undergraduate students in a university;
- The course should not presume prior knowledge of another programming language;
- The students should be expected to continue to study a subsequent course in computer programming;
- The outcomes of the study should include at least one that measures individual achievement at the end of the module/unit;
- The improvement should not be attributable to the removal of a technical constraint that would no longer apply;
- The intervention should introduce some additional use of ICT.

Criterion 1 excludes training courses, where the context of study would change the motivation of students—such as Carlisle (2010), and/or those that require students to perform at a higher level of abstraction. Criterion 2 excludes experiments where subjects may have a prior mental model. Criterion 3 ensures that an identifiable subgroup of the students are motivated to view this stage as a starting point, and excludes studies where learners may only have short-term learning objectives (see, Nagappan et al. 2003, and elsewhere). Studies that could not be generalized to a ‘standard’ cohort would also be excluded, as in Katai and Toth (2010). Criterion 4 ensures that the change in practice is measurable within the current assessment framework for higher education in this field. This eliminates studies that are over-focused on immediate or short-term gains (e.g. Rajala, Laakso, Kaila, and Salalkoski 2008, or Chen 2011), those that are over-designed to the intervention (e.g. de Kereki, Azpiazu & Silva, 2004), or those that would not have a direct impact on assessment, as in Mow, Au and Yates (2006). Criterion 5 ensures that any intervention would still produce comparable improvements in the current technological environment—see, for example, Lidtke (1979). Criterion 6 limits the study to those that depend on educational technology.

Although it is conventional in medicine to provide an explicit definition for these criteria, experience from this study suggests that this is far from easy. Variation in the structural organisation of higher education courses is sufficiently diverse to have required frequent redrafting of the criteria in order to ensure that studies that were clearly relevant were not excluded by errors in drafting the criteria. Criterion 3 proved particularly susceptible: some universities require that a student makes a choice of a major at the start of the first year, some allow the decision to be optional, whilst delay that choice (McDowell, Werner, Bullock, and Fernald 2003). However, as the review process records the criteria on which a study is rejected, it is always possible to apply any revision of the criteria retrospectively.

Any study that fails this stage will fail at least one criterion on the basis of information that was explicitly stated in one or more of the papers describing the study. Oh (1988) provides an illustrative example. This study of cooperative learning focused on seven weeks of a course that consisted of 106 first and second year students, of whom business majors provided the largest sub-group. The course was a 'short-course' for the business majors and credit for the course could not be counted towards a final degree. There is no indication that the subsequent stages of the course required further study of programming, and no details were presented that could be used to separate the performance of different sub-groups. This was sufficient to exclude the study on the criterion 3. A more detailed reading reveals that within groups, 50% had previous experience with a previous programming language. Although it might have been possible to consider exclusion on conditions 2 and 4 as well, failure on criterion 3 is sufficient to exclude the study from subsequent analysis.

One further criterion, that the study should include at least 100 subjects, was drafted to ensure that any beneficial effects would be transferable to large cohorts; this was discarded in the early stages. Issues of scaling relate directly to the differences between effect, efficacy, effectiveness and efficiency (Alsop & Tompsett 2007). Large-scale experiments might include only a few relevant students (e.g., Awad, 1989), and interventions can be trialled using a small, but correctly selected sample from a large cohort. Some interventions could be scaled up to large cohorts without a significant increase in resources, but this is not always so. As the applicability of any intervention depends on both the cost of resources and the benefit of any increase in learning and the latter can only be assessed after all studies of the same intervention have been assessed.

## The studies that remain

Despite the number of studies that were reviewed, only four studies remained after the inclusion/exclusion phase. Each of these studies provided sufficient contextual information to compare the outcomes of an intervention with one group of students, with the outcomes of those from another group, and present sufficient data for compare the groups. Table 2, below, provides summary information for these four studies indicating the innovation that was introduced. No consideration of the magnitude, direction and/or significance of any effect is ignored until it is clear that these can be associated with the intervention that was introduced.

*Table 2. Studies providing relevant evidence for the introduction of additional ICT.*

Year <sup>1</sup>	Key Reference	Intervention
1997	Foubister, Michaelson & Tones (1997)	Introduction of automatic submission/feedback system
1999	Catenazzi & Sommaruga (1999)	Use of Hyper Apuntes learning environment
2004	Daly & Horgan (2004)	Use of RoboProf programming environment
2008	Smith & Fidge (2008)	Optional use of multimedia resources

<sup>1</sup>Year of study

In each case there was sufficient information regarding the research process to provide a 'nominal' classification of the study at a specific level of evidence relevant to the study, such as:

- a randomised controlled trial (RCT—eliminates bias from unknown variables),
- a cohort study (non-random allocation, prospective, with follow through), or
- a case-control comparison (retrospective, but potentially biased towards intervention)

Studies at level 4 (case series), would not be relevant as it is not possible to separate effects that arise from the intervention and those that result from the selection of the participants.

## Level of evidence provided for an intervention

The actual level of evidence in each of these studies was determined in two stages. In the first stage, each study was assessed in terms of the process that was used to allocate students to the experimental and/or control/reference groups. In the second stage, the level is reviewed for uncontrolled sources of bias. It would have been reassuring to find the descriptions of each study included the relevant information to so, but this information was often hard to determine.

Randomized allocation of subjects to a control and intervention group is an essential element in any level 1 study, but it is difficult to achieve this in educational studies. Only Catenazzi and Sommaruga (1999) applied randomization, but they did not establish the preconditions for an RCT. In this study the treatment group was selected randomly from those who wished to participate (40/100), but the control group was composed of those who did not wish to participate in the experiment, rather than from those who were not selected—ensuring that treatment and control groups were non-comparable. This removes the conditions that are necessary to draw any conclusions regarding differences in outcome between the two groups, and limits the level of evidence for this study to a *case series* (level 4) at best.

In the remaining three studies, the reference group is taken as a previous cohort and is therefore retrospective in design. This reduces the level of evidence to a case-control study (level 3, or below), and differences in outcome are dependent on both the intervention and other differences between the student populations in each year. The study by Foubister, Michaelson and Tomes (1997) nominally meets the criteria for this level. In the study by Smith and Fidge (2008), the change in year between the two groups was associated, in addition, with a change in both the curriculum and the assessment, and they acknowledge this as a limit on their findings. This study must also be classified as a *case series* (level 4). Daly and Horgan's study from 2004, is restricted to a single cohort, and optionality was used to differentiate the intervention and reference groups. Without further analysis to show that the groups could be treated as equivalent, the results must also be treated as a *case series*.

Table 3, below, shows these studies in summary form, sorted in terms of quality of evidence, level of significance claimed for the results, and alphabetically.

*Table 3. Level of evidence for remaining studies*

Level <sup>1</sup>	Key Reference	Significance <sup>2</sup>
3	Foubister, Michaelson and Tomes (1997)	0.29
4	Catenazzi and Sommaruga (1999)	none claimable
4	Daly and Horgan (2004)	none claimable
4	Smith and Fidge (2008)	none claimed

<sup>1</sup>Level of evidence

<sup>2</sup>Probability that the intervention had no measurable effect

It is clear that none of these studies demonstrates a statistically significant change in performance as a direct result of the intervention that was introduced, and further analysis (as a systematic review) is impossible. Despite the significant number of studies that were reviewed, only one study establishes the basis for a comparison of relevant learning outcomes between two groups, where one group has had the benefit of additional ICT, and the other has not, and in this study, the results are not significant. As an overall summary, it would appear that none of the research published in English that has addressed this educational problem can provide evidence for a change in teaching practice.

## Conclusions

The intention of this review was to provide a quasi-systematic review of all the innovative applications of ICT that have been designed to improve learning of a first computer language on a computer science course in higher education. Following Alsop and Tompsett (2007), it was anticipated that a reasonable number of studies would be identified, the interventions could be ranked according to the level of evidence (and phase of research), and that the majority would be at level A(2)—educational effect. In this domain, the outcome is even more surprising: if effect is measured in terms relevant to assessment in higher education, then the only intervention supported by research at a level that warrants generalization is for an automatic submission/feedback system, and, even for that innovation, the difference in outcomes was not significant.

Without any reliable and measurable educational benefit for an ICT-based intervention, the rational approach for practitioners in this domain, would be to consider non-ICT-based innovations (as suggested above), other non-ICT based approaches (as already discussed in the section on possible changes), or to discard innovation and evolve their own approach to correspond to the uniqueness of their own educational practice.

## Discussion

The fundamental factor that has led to this conclusion is clearly the small number of studies that remained, as determined by the criteria, and it is tempting to suggest that a more positive analysis would have resulted from less stringent criteria. This is not the case, however. Most studies were eliminated from the analysis because they did not establish sufficient grounds to compare the outcomes of a relevant subgroup of students with a comparable reference group. On the basis of the papers, many studies appear to have included relevant students but did not distinguish these students within the experimental design. Other studies identified relevant sub-groups of students, but failed to account for these sub-groups within the subsequent analysis. In any of these cases a 'better-designed' experiment, or more appropriate analysis, would have allowed the experiment to remain within the study. Even where the analysis allowed relevant sub-groups to be compared, the majority failed to demonstrate that any measured difference in outcome was independent of the influence of variables that are already associated with differences in learning in this domain at this level. In such studies, any effect would be limited to an equivalent context—without providing the information that is required to establish equivalence. In these cases, it might be sensible for authors to recognize this in their own conclusions, and repeat the experiment and include the relevant information, but few do so. If 'technicians' expect practitioners to change, then research must be designed so that the results are transferable (irrespective of the methodology used).

Others might wish to argue that higher standards of 'design' cannot be achieved in any educational field, but this is contradicted by other experimental work in education. At the same domain and level, a study by Beck, Chizhik and McElroy in 2005, introducing group work as a change in the pattern of study (without additional use of ICT), was completed to a higher standard than any of those considered here, and a successful sequence of experiments following a design experiment methodology (Collins, 1999) have progressed beyond measuring educational effect at different levels in other domains (see, for example, Collins, Joseph and Bielaczyc, 2004), taking into account a wide range of factors that can affect the introduction of innovative practice at local level. Even complex problems, such as the effect of class-size on pupils' achievement, have been successfully researched using 'natural experiments' (Angrist & Levy, 1999). None of these studies came within the scope of this review.

A less palatable explanation, but one that would reflect the lack of progress over 45 years of technological development, was suggested by Alsop and Tompsett's paper in 2007: it is simpler to conduct and publish research to demonstrate that new technology has some effect, rather than to conduct well-designed experiments into the educational benefits of established technological interventions, if the latter are less likely to produce novel findings.

## Limitations

A number of factors must be seen as limiting the validity of this conclusion, even if they are normally accepted without question, in other reviews. Firstly, the studies that have been considered were limited to those published in English. It might be tempting to suggest that key advances published in another language would also be republished in English, but this is difficult to justify. It is likely that far more students are regularly taught a first programming language in a different language, and the republication of such research in English is far from assured. Secondly, although the search and classification processes are fully traceable, it is possible that a panel of co-researchers might have identified a wider set of papers in the first instance. In this respect this study cannot be classified as a full systematic review. Whilst that remains a possibility, the resources that would be required to conduct the study would have exceeded, by far, those who are searching for relevant ICT-based interventions within this field.

A full listing of all the papers and studies is available, on request, to the author.

## References

- Alsop, G., & Tompsett, C. (2007). From effect to effectiveness the missing research questions. *Education, Technology and Society*, 10(1), 28–39.
- Angrist, J. A., & Levy, V. (1999) Using maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics*. 114(2), 533–575.

- Awad, A. M. (1989) *Analysis of college students' behavior and characteristics in the process of learning computer programming*. (Unpublished doctoral thesis). School of Graduate Studies, University of Missouri, Kansas City.
- Bayman, P. (1983) Novice users' misconceptions of basic programming statements (Technical report no. 82-1). Santa Barbara, CA: California University.
- Bayman, P., & Mayer R. E. (1988). Using conceptual models to teach basic computer programming. *Journal of Educational Psychology*, 80(3), 291–298.
- Beck, L. L., Chizhik, A. W., & McElroy A. C. (2005). Cooperative learning techniques in CS1: Design and experimental evaluation. *Proceedings of the 36th SIGCSE technical symposium on Computer science education* (pp. 470–474). New York, USA: ACM.
- Bloch, S. A. (2000). SCHEME and JAVA in the first year. In J. G. Meinke (Ed.), *Fifth Annual CCSC Northeastern Conference on the Journal of Computing in Small Colleges 2000*, Consortium for Computing Sciences in Colleges (pp. 157–165). Mawash, NJ: Consortium for Computing Sciences in Colleges.
- Bornat R., Dehnadi, S., & Simon. (2010). Mental models, consistency and programming aptitude. In S. Hamilton & M. Hamilton (Eds.), *Tenth Conference on Australasian Computing Education* (January 2008), Australian Computer Society, 78, 53–61. Darlinghurst, Australia: Australian Computer Society.
- Boyer, K. E., Lahti, W., Phillips, R., Wallis, M. D., Vouk M. A., & Lester J. C. (2010). Principles of asking effective questions during student problem solving. In Lewandowski et al. (Eds.), *Proceedings of the 41st ACM technical symposium on Computer science education* (pp. 460–464). New York, USA: ACM.
- Brusilovsky, P., Calabrese, E., Hvorecky, J., Kouchnirenko, A., & Miller, P. (1997). Mini-languages: A way to learn programming principles. *Education and Information Technologies* 2(1), 65–83.
- Burnett, J. (1979). Self-paced Fortran. *Educational Research and Methods*, 11(2), 53–56.
- Campbell, P. F., & McCabe, G. P. (1984). Predicting the success of freshmen in a computer science major. *Communications of the ACM*, 27(11), 1108–1113.
- Carlisle, M. C. (2010). Using YouTube to enhance student class preparation in an introductory JAVA course. In G. Lewandowski et al. (Eds.), *Proceedings of the 41st ACM technical symposium on computer science education* (pp. 470–474). New York, USA: ACM.
- Carlisle, M. C., Wilson, T. A., Humphries, J. W., & Hadfield, S. M. (2005). RAPTOR: A visual programming environment for teaching algorithmic problem solving. *Proceedings of 36th Technical Symposium on Computer Science Education* (pp. 176–180). New York, NY: ACM.
- Catenazzi, N., & Sommaruga, L. (1999). The evaluation of the HYPER-APUNTES interactive learning environment. *Computers and Education*, 32(1), 35–49.
- Checkland, P. (1999). *Soft systems methodology: A 30-year retrospective*. Wiley Chichester, UK: Chichester.
- Chen, J. W. & Lin, C.-C. (2006). A van-Hiele web-based learning system with knowledge management for teaching programming. In Kinshuk et al. (Eds.), *Proceedings of the Sixth IEEE International Conference on Advanced Learning Technologies* (pp. 114–116). Piscataway, NJ: IEEE Press.
- Chen, L.-H. (2011) Enhancement of student learning performance using personalized diagnosis and remedial learning system. *Computer Education*, 56(1), 289–299.
- Cheney, P. H. (1977) Teaching computer programming in an environment where collaboration is required. *AEDS Journal*, 1(1), 1–5.
- Collins, A. (1999). The changing infrastructure of education research. In E. C. Langemann & L. S. Schulman (Eds.), *Issues in education research* (pp. 15–22). San Francisco, CA: Jossey-Bass
- Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *Journal of the Learning Sciences*, 13(1), 15–42.
- Cooper, S., Dann, W., & Pausch, R. (2000). ALICE: A 3-D tool for introductory programming concepts. *Journal of Computing Sciences in Colleges*, 15(5), 107–116.
- Corney, M., Teague, D., & Thomas, R. N. (2010) Engaging students in programming. In T. Clear & J. Hamer, *Proceedings of the Twelfth Australasian Conference on Computing Education* (pp. 63–72). Darlinghurst, Australia: Australian Computer Society.

- Crawford, S., & Boese, E. (2006). ACTIONSCRIPT: A gentle introduction to programming. *Journal of Computing Sciences in Colleges*, 21(3), 156–168.
- Daly, C., & Horgan, J. M. (2004). An automated learning system for java programming. *IEEE transactions on Education*, 47(1), 383–387.
- de Kereki, I. F., Azpiazu, J., & Silva, A. (2004). Knowledge management in learning environment design. *34th Frontiers in Education*, 3, S3F6–11. doi: 10.1109/FIE.2004.1408777
- de Raadt, M., Hamilton, M., Lister, R., Tutty, J., Baker, B., Cutts, Q., ...Tolhurst, D. (2005, July). *Approaches to learning in computer programming students and their effect on success*. Paper presented at the Annual International Conference of the Higher Education Research and Development Society of Australasia, Milperra, NSW, Australia.
- Emurian, H. H., Hu, X., Wang, J., & Durham, A. G. (2000). Learning JAVA: A programmed instruction approach using applets. *Computers in Human Behavior*, 16(4), 395–422.
- Falkner, K., Palmer, E. (2009). Developing authentic problem solving skills in introductory computing classes. *SIGCSE Bulletin*, 41(1), 4–8.
- Fletcher, S. H. (1984). *Cognitive abilities and computer programming*. Retrieved from ERIC database. (ED259700).
- Ford, M., & Venema, S. (2010). Assessing the success of an introductory programming course. *Journal of Information Technology Education*, 9, 133–145.
- Foubister, S. P., Michaelson, G., & Tomes, N. (1997). Automatic assessment of elementary standard ML programs using CEILIDH. *Journal of Computer Assisted Learning*, 13(2), 99–108.
- Golding, P., Donaldson, O., & Tennant, V. (2009). Application of modified perceived learning problem inventory (PLPI) to investigate performance in introductory programming. In J. P. T. Higgins & S. Green, *Proceedings of the 39th IEEE international conference on Frontiers in education conference* (pp. 366–371). Piscataway, NJ: IEEE Press.
- Hagan, D., & Macdonald, I. (2000). A collaborative project to improve teaching and learning in first year programming. *Australasian Journal of Engineering Education*, 9(1), 65–76.
- Hammersley, M. (2002). *Educational research policymaking and practice*. London, UK: Paul Chapman.
- Hargreaves, D. H. (1996). *Teaching as a research-based profession: possibilities and prospects*. London, UK: Teacher Training Agency.
- Higgins, J. P. T., & Green, S. (2009). Guide to the contents of a Cochrane protocol and review. In J. P. T. Higgins & S. Green, *Cochrane Handbook for Systematic Reviews of Interventions* (pp. 52–79). Chichester, England: John Wiley.
- Jadad, A. R. (1998) *Randomised controlled trials*. London, UK: BMJ Books.
- Johnson, J. A., & Johnson, G. M. (1992). Student characteristics and computer programming competency: A correlational analysis. *Journal of Studies in Technical Careers*, 14(1), 33–46.
- Johnson, W. L., & Soloway, E. (1985). PROUST. Knowledge-based program understanding. *IEEE Transactions on Software Engineering*, 11(3), 267–275.
- Katai, Z., & Toth, L. (2010). Technologically and artistically enhanced multi-sensory computer programming education. *Teaching and Teacher Education*, 26(2), 244–251.
- Katz, S., Aronis, J., Allbritton, D., Wilson, C., & Soffa, M. L. (2003). A study to identify predictors of achievement in an introductory computer science course. *Proceedings of the SIGMIS Conference on Computer Personnel Research: Freedom in Philadelphia-leveraging differences and diversity in the IT workforce* (pp. 157–161). New York, NY: ACM.
- Lewis, T.L., Chase, J. D., Pérez-Quinones, M.A., & Rosson, M. B. (2005). The effects of individual differences on cs2 course performance across universities. *Proceedings of the 36th SIGCSE Technical Symposium on Computer Science Education* (pp. 426–430). New York, NY: ACM Press.
- Li, C, Wang, H., & Zhang, H. (2006). A case of blending learning in computer teaching—the model and the application. In R. Mizoguchi, P. Dillenbourg, & Z. Zhu (Eds.), *Proceeding of the 2006 conference on Learning by Effective Utilization of Technologies: Facilitating Intercultural Understanding* (pp. 45–48). Amsterdam, The Netherlands: IOS Press.
- Lidtko D. K. (1979). *Paired and individual learning in computer programming: An evaluation of student achievement and cost effectiveness*. (Unpublished doctoral thesis). University of Oregon, Oregon.

- Lou, Y., Abrami, P. C., & d'Apollonia, S. (1996). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research*, 71(3), 459–521.
- Mayer, R. E. (1975). *Instructional variables in computer programming: Final report*. Bloomington, IN: Indiana Mathematical Psychology Program.
- McDowell, C., Werner, L., Bullock, H. E., & Fernald, J. (2003). The impact of pair programming on student performance, perception and persistence in proceedings. *Proceeding of the 25th International Conference on Software Engineering* (pp. 602–607). Washington, DC: IEEE Computer Society.
- Mow, I. C., Au, W. K., & Yates, G. C. R. (2006). The impact of cable on teaching computer programming. In R. Mizoguchi, P. Dillenbourg, & Z. Zhu (Eds.), *International Conference on Computers in Education* (pp. 55–62). Amsterdam, The Netherlands: IOS Press.
- Nagappan, N., Williams, L., Ferzli, M., Wiebe, E., Yang, K., Miller C., & Balik S. (2003). Improving the CS1 experience with pair programming. *Proceedings of the 34th SIGCSE technical symposium on Computer Science Education* (pp. 359–362). New York, New York, NY: ACM.
- Nowaczyk, R. H. (1984). The relationship of problem-solving ability and course performance among novice programmers. *International Journal of Man-Machine Studies*, 21(2), 149–160.
- Oancea, A. (2005). Criticisms of educational research: Key topics and levels of analysis. *British Educational Research Journal*, 31 (2), 157–183.
- OECD CERI (2002). *Educational research and development in England*. Paris, France: OECD Publishing.
- Oh, H.-A. (1988). *The effects of individualistic, cooperative task, cooperative incentive structures on college student achievement in computer programming in BASIC*. (Unpublished doctoral thesis). Department of Curriculum and Instruction, Illinois State University, Illinois.
- Ophel, J., Prosser, M., & Robinson, I. (1997, July). *First year university students' experiences of studying computer science*. Paper presented at the Conference on Learning and Teaching in Higher Education, Adelaide, Australia.
- Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire. *Educational and Psychological Measurement*, 53(3), 801–813.
- Rajala, T., Laakso, M.-J., Kaila, E., & Salalkoski, T. (2008). Effectiveness of program visualization: A case study with the Ville tool. *Journal of Information Technology Education*, 7, 15–32.
- Reid, K. L., & Wilson, G. V. (2005) Learning by doing: introducing version control as a way to manage student assignments. *Proceedings of the 36th SIGCSE Technical Symposium on Computer Science Education* (pp. 272–276). New York, NY: ACM Press.
- Rogers, E. M. (1962). *Diffusion of Innovations*. New York, NY: The Free Press.
- Sajaniemi, J., Kuittinen, M., & Tikansola, T. (2008). A study of the development of students' visualizations of a program state during an elementary object-oriented programming course. *Journal on Educational Resources in Computing*, 7(4), 1–31.
- Serbec, I. N., Kaucic, B., & Rugelj, J. (2008). Pair programming as a modern method of teaching computer science. *International Journal of Emerging Technologies in Learning*, 3, 45–49.
- Sheard, J., Carbone, A., & Markham, S. (2000, November). *The process of learning programming: A comparative study of students' reactions*. Paper presented at the Eighth International Conference on Computers in Education/ International Conference on Computer-Assisted Instruction: Learning Societies in the New Millennium: Creativity, Caring and Commitments, Hsinchu, Taiwan.
- Sheard, J., Simon, S., Hamilton, M., & Lönnberg, J. (2009) Analysis of research into the teaching and learning of programming. In M. J. Clancy, M. E. Caspersen, & R. Lister (Eds.), *Proceedings of the fifth international workshop on Computing education research workshop* (pp. 93–104). New York, NY: ACM.
- Simons, H. (2003). Evidence-based practice: panacea or over promise? *Research Papers in Education*, 18(4), 303–311.
- Smith, G., & Fidge, C. (2008). On the efficacy of prerecorded lectures for teaching introductory programming. In S. Hamilton & M. Hamilton (Eds.), *Tenth Australasian Computing Education Conference on Research and Practice in Information Technology* (pp. 129–30). Darlinghurst, Australia: Australian Computer Society.
- Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research*, 69(1), 21–51.

- Tan, P.-H., Ting, C.-Y., & Ling, S. W. (2009). Learning difficulties in programming courses: Under- graduates' perspective and perception. *Proceedings of the 2009 International Conference on Computer Technology and Development* (Vol. 1, pp. 42–46). Los Alamitos, CA: IEEE Computer Society.
- Tew, A., Dorn, B., Leahy, W. D. J., & Guzdial, M. (2008). Context as support for learning computer organization. *Journal on Educational Resources in Computing*, 8(3). doi: 10.1145/1404935.1404937
- Thornburg, K. M. (2010). *The effect of positive and negative messages on problem solving in computer programming tasks*. (Unpublished doctoral thesis). The Graduate College, University of Iowa, Iowa City.
- Thweatt, M. (1994). CS1 closed-lab vs. open-lab experiment. *Proceedings of the Twenty-Fifth SIGCSE Symposium on Computer Science Education (SIGCSE '94)* (pp. 80-82). New York, NY: ACM.
- Truong, N., Bancroft, P., & Roe, P. (2003). A web based environment for learning to program. In Oudshorn, M. J. (Ed.), *Proceedings of the 26<sup>th</sup> Australasian Conference on Computer Science Education* (pp. 255–264). Darlinghurst, Australia: Australian Computer Society.
- Williams, L., Wiebe, E., Yang, K., Ferzli, M., & Miller, C. (2002). In support of pair programming in the introductory computer science course. *Computer Science Education*, 12(3), 197–212.



## A Review of Intervention Studies On Technology-assisted Instruction From 2005-2010

**Ying-Tien Wu<sup>1</sup>, Huei-Tse Hou<sup>2</sup>, Fu-Kwun Hwang<sup>3</sup>, Min-Hsien Lee<sup>4</sup>, Chih-Hung Lai<sup>5</sup>, Guo-Li Chiou<sup>6</sup>, Silvia Wen-Yu Lee<sup>7</sup>, Yu-Chen Hsu<sup>2</sup>, Jyh-Chong Liang<sup>2</sup>, Nian-Shing Chen<sup>8</sup> and Chin-Chung Tsai<sup>9\*</sup>**

<sup>1</sup>Graduate Institute of Network Learning Technology, National Central University, Taiwan // <sup>2</sup>Graduate Institute of Applied Science and Technology, National Taiwan University of Science and Technology, Taiwan // <sup>3</sup>Department of Physics, National Taiwan Normal University, Taiwan // <sup>4</sup>Center for Teacher Education& Institute of Education, National Sun Yat-sen University, Taiwan // <sup>5</sup>Department of Computer Science and Information Engineering, National Dong Hwa University, Taiwan // <sup>6</sup>Institute of Education, National Chiao Tung University, Taiwan // <sup>7</sup>Graduate Institute of Science Education, National Changhua University of Education, Taiwan // <sup>8</sup>Information Management Department, National Sun Yat-sen University, Taiwan // <sup>9</sup>Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taiwan // ytwu@cl.ncu.edu.tw // hthou@mail.ntust.edu.tw // hwang@phy.ntnu.edu.tw // leemh@mail.nsysu.edu.tw // laich@mail.ndhu.edu.tw // glchiou@mail.nctu.edu.tw // silviawyl@cc.ncue.edu.tw // hsuyuchen@mail.ntust.edu.tw // aljc@mail.ntust.edu.tw // nschen@mis.nsysu.edu.tw // cctsai@mail.ntust.edu.tw //

\*Corresponding author

(Submitted April 12, 2012; Revised June 29, 2012; Accepted August 20, 2012)

### ABSTRACT

By reviewing papers published in five important SSCI journals from 2005 to 2010, this study aimed to provide insights into intervention studies on technology-assisted instruction. It was found that still relatively little research on educational technology which addresses the effects of specific instructional interventions on student learning was conducted during this period. Moreover, most reviewed studies were conducted in higher education, rather than in high school, elementary school, or adult education contexts. The two subject domains, science and engineering (including computer studies), were most frequently involved in these studies, the majority of which addressed achievement as research (learning outcome) foci, while relatively fewer studies investigated students' learning process or affective outcomes. Regarding technology adoption, this study revealed that, in both the 2005-2007 and 2008-2010 periods, technologies for specific instructional purposes(e.g., a specially designed system for online collaborative writing) were more frequently adopted than those for general purposes (such as PowerPoint). This study also reveals that technology-mediated interpersonal interactions are commonly utilized in these intervention studies, with the focus mainly on student-student interactions.

### Keywords

Technology-assisted instruction, Intervention studies, Research trends

### Introduction

In the past two decades, advancements in information and communication technology have made considerable impacts on educational practices. In particular, the use of technology in assisting teaching and learning has become a new educational paradigm, and teachers are largely advised to adopt technology more to enrich their instructional practices (Lawless & Pellegrino, 2007).

Many educators and researchers have advocated that pedagogical considerations are crucial in the use of technology in education (e.g., Leijen et al., 2008). From their perspectives, instructional designs within technology-assisted instruction should be highlighted. In the past, a large body of research addressing the use of technology in instructional practices has been undertaken. These studies focused on various issues and were conducted using different research methods. In particular, some of them were conducted with specific instructional interventions. An intervention study means that certain technology-assisted instructional activities are conducted and empirical data from participants are collected for analysis and evaluation. Reviewing intervention studies published in the literature can help us to understand the actual influences of using technology in instructional practices.

In recent years, some reviews and meta-analyses focused on issues regarding technology-assisted instruction have been conducted (e.g., Bernard et al., 2004; Bernard et al., 2009; Dillon & Gabbard, 1998; Fabos & Young, 1999; Lawless & Pellegrino, 2007; Lee et al., 2011; Straub, 2009; Tamim et al., 2011; van Rooij, 2009). These review studies had various foci, including comparisons of the instructional effects between technology-assisted instruction and traditional instruction (e.g., Bernard et al., 2004; Tamim et al., 2011), meta-analysis of the interactive design of technology-assisted instruction (Bernard et al., 2009), and reviews of the use of specific technologies in instruction (Dillon & Gabbard, 1998; Fabos & Young, 1999; van Rooij, 2009), as well as reviews of teachers' professional developments in technology-assisted instruction (Lawless & Pellegrino, 2007; van Rooij, 2009). However, it should be noted that most of these meta-analyses merely addressed the effects of certain specific forms of technology-assisted instruction, such as hypermedia (Dillon & Gabbard, 1998), telecommunication (Fabos & Young, 1999), and open-source software (van Rooij, 2009). None of them were conducted to provide an overall understanding of intervention studies on technology-assisted instruction. As an intervention study on technology-assisted instruction is conducted in a real learning environment (physical or online) with certain research treatments, this kind of study can provide educators and instructors with more authentic and meaningful information regarding the effectiveness of the specific technology-assisted instruction. An overall review of intervention studies on technology-assisted instructions would be helpful in providing important insights into the future research tendencies. To this end, this study was conducted to review intervention studies on technology-assisted instruction published in the literature from 2005 to 2010.

Besides, most of the existing reviews or meta-analyses in technology-assisted instruction merely analyzed its effects on learners' achievements (e.g., Bernard et al., 2004; Bernard et al., 2009; Tamim et al., 2011). Recently, more and more relevant studies have focused on the effects of technology-assisted instruction on learners' affective learning outcomes, such as attitudes, motivation, and learning processes (e.g., Hewitt, 2003, 2005; Hou & Wu, 2011; Lee & Tsai, 2011; Wever et al., 2006). However, a comprehensive analysis of all kinds of research foci (i.e., including learning achievements, affective learning outcomes and process) based on intervention studies of technology-assisted instruction is still not available. Therefore, the research foci of the reviewed intervention studies on technology-assisted instruction from 2005-2010 were analyzed in this study.

Generally speaking, technology-assisted instruction involves technology adoption of various software or hardware. The software or hardware used in intervention studies on technology-assisted instruction may have been developed for general and widespread usages, such as PowerPoint (Susskind, 2008), or for specific instructional purposes, such as learning systems or educational games. Therefore, the technology adoption (adopted for general or specific purposes) in the reviewed intervention studies on technology-assisted instruction (i.e., technologies used in these studies are for general purposes or for specific instructional purposes) was analyzed in this study.

In addition, the rapid development of information technologies in Web 2.0 has made interpersonal interaction in learning environments more flexible and efficient. As a result, various novel information technologies or tools have been commonly used in technology-assisted instruction to promote interpersonal interactions. Therefore, interpersonal interactions, including interaction types (i.e., face-to-face interaction, technology-mediated interaction, and blended interaction) and participant interactions (i.e., student-teacher, student-student, and both), were analyzed in this study.

Finally, some cross-analyses were also conducted among technology adoption, sample group and subject domain. Linn (2003) proposed that new technologies generally support user customization, indicating that while adopting technologies to assist learning or teaching, the features of subject domains and characteristics of learners should be taken into account. Therefore, this study further conducted cross-analyses between the technology adoption and the sample groups (as well as subject domains). Moreover, the use of technologies may have changed the manner of interpersonal interactions. To further understand the possible associations between interpersonal interaction types and participant interactions, cross-analyses between interaction types and participant interactions were also conducted.

To sum up, this study aimed to conduct a meta-analysis review of the empirical intervention studies on technology-assisted instruction published in the literature for a period of six years from 2005 to 2010. The research purposes of this study are:

- To explore the future research trends in sample groups, subject domains, research foci, technology adoption, and interpersonal interaction (including interpersonal interaction types and participant interactions) based on intervention studies on technology-assisted instruction articles published from 2005 to 2010.

- To investigate how the research trends revealed from the reviewed studies may differ in terms of sample groups, subject domains, research foci, technology adoption, and interpersonal interaction from 2005 to 2007 and 2008 to 2010.
- To conduct cross analyses among the sample groups, subject domains, interpersonal interaction, and technology adoption for the reviewed intervention studies on technology-assisted instruction.

## Method

### Research papers for analysis

This study aimed to review the intervention studies on technology-assisted instruction published in 2005-2010. To this end, the papers published in the five Social Sciences Citation Index (SSCI) journals on educational technology, including the *British Journal of Educational Technology* (BJET), *Computers & Education* (C&E), *Educational Technology Research and Development* (ETRD), the *Journal of Computer Assisted Learning* (JCAL), and the *Journal of Educational Technology & Society* (ET&S), were selected as the literature source for this review. There were a total of 4,093 papers published in these five journals from 2005 to 2010.

### Review processes

There were two stages of the review conducted in this study. In the first stage, all of the empirical intervention studies on technology-assisted instruction published in the five journals from 2005-2010 were selected. In this stage, three individual researchers read the full-text of all the publication items and systematically screened the articles to identify those to be analyzed. There were two criteria used for the paper selection in this stage. First, the study should include specific technology-assisted instruction as intervention for designing activities. Second, the study should collect empirical data with at least one of the following types: achievement, process, or attitudes from the designed learning activities. It should be noted that those studies simply describing the design of educational technology or a learning system without the support of empirical data were excluded from this review. The three individual researchers with doctoral degrees or professorship in the instructional technology field conducted the screening tasks and had discussions in order to reach overall consensus based on the described selection rules. The experts read the full text of each article one by one to decide if it could be included in this review. Finally, a total of 322 articles out of 4,093 were selected for the analyses in this study.

In the second stage, a coding scheme (a detailed description is given below) was discussed and established by the authors. Then, a series of content analyses were conducted by using the articles selected in the first stage. All the reviewers (i.e., the authors) in this study have doctoral degrees or professorship in the instructional technology field, and most of them have published more than one article in the five key journals. After the two stages of the review in this study, all the reviewers (i.e., the authors) discussed and reached consensus on the interpretations of the quantitative results in this study.

### Coding scheme

The coding scheme used in this study consists of six major categories. Each category contains several sub-categories. A detailed description of the categories is provided below:

### Sample groups

In this study, four sample groups were identified based on the educational backgrounds of the participants in all the reviewed studies. These groups are: (1) elementary school students (K-6 graders), (2) high school students (7-12 graders), (3) higher education students (college and graduate students), and (4) adults. If a study had more than one sample group, only one major sample group was coded.

## Subject domain

Similar to Hsu et al. (2012), the subject domains in this study were grouped into seven sub-categories: (1) Science (e.g., Physics, Chemistry, Biology, Medical and Sport Science), (2) Mathematics, (3) Arts & Language, (4) Social Studies, (5) Engineering (including Computers), (6) Others, and (7) Unclear. If a study involved more than one subject domain, only one major subject domain was coded.

## Research focus

In this study, the research foci (learning outcome) of all the selected studies were divided into the three major categories: *achievement* (Ac), *learning processes* (Pr), and *affective outcomes* (Af). A study with a research focus on “*achievement*” aims to explore learners’ recall of information or acquisition of knowledge or skills measured by test scores (e.g., Truman & Truman, 2006) or by other assessments, such as concept maps (Hoskins & van Hooff, 2005); a research focus on “*learning process*” may investigate, for example, the patterns of online discussion (e.g., Liu & Tsai, 2008), learning approaches used by students (e.g., Fessakis, Tatsis, & Dimitracopoulou, 2008), or interactions between students (e.g., Monteserin, Schiaffino, & Amandi, 2010); a study addressing “*affective outcomes*” may explore perceptions of the subject domain, perceptions of the learning, peer acceptance (e.g., Balram & Dragičević, 2008), enjoyment of the learning materials (e.g., Grimshaw, Dungworth, McKnight, & Morris, 2007), or satisfaction with the course (e.g., Alonso, Manrique, & Viñes, 2009). It should be noted that a reviewed study may contain more than one research focus. According to its multiple research foci, the “research focus” of such a study was coded as “Ac, Pr,” “Ac, Af,” “Pr, Af,” or “Ac, Pr, Af.”

## Technology adoption

To explore the issue of technology adoption, each of the reviewed studies was examined to identify its adopted technology and was categorized as either *technology for general purpose* or *technology for specific instructional purpose* for its technology adoption attribute. For example, a study identified as using technologies for general purposes means the software and/or hardware used had been developed for common and widespread usages, such as PowerPoint, Wikipedia, E-mail or asynchronous discussion forums. For example, in Susskind (2008), the PowerPoint software was used, and the effects of accompanying lectures with computer-mediated PowerPoint presentations or PowerPoint generated overheads on students’ self-efficacy, attitudes, course performance, and class-related behaviors were examined. On the other hand, a study identified as using technologies for a specific purpose means the software/hardware was designed for specific instructional activities and needs, such as a specially designed course management system, a learning management system or a virtual learning environment. For instance, in Cho and Schunn (2007), a web-based reciprocal peer review system for scaffolding writing and rewriting was specially developed to help students generate constructive comments on others’ writing and to facilitate the administration of students’ writings and reviews. Therefore, the technology adoption of Cho and Schunn (2007) was coded as “technology for specific instructional purpose” in this study.

## Interpersonal interaction type

According to the face-to-face interaction and technology-mediated interaction used, the interaction type of each study was coded into one of three categories: *face-to-face interaction*, *technology-mediated interaction*, and *blended interaction*. Technology assisted face-to-face interaction means that teachers and students interacted face to face with the assistance of technologies. For example, a teacher conducted face-to-face instruction while using an electronic whiteboard (e.g., Lopez, 2010) or students discussed together in face-to-face mode via a shared screen with digital contents (e.g., Zydny, 2010). In terms of technology-mediated interaction, technologies are used as the major channel for facilitating interactions, such as a web-based discussion forum (e.g., Wei & Chen, 2006). As for blended interaction, it refers to using both face-to-face and technology-mediated interactions for instruction (e.g., Looi, Chen, & Ng, 2010).

## Participant interaction

In this study, the participant interaction is defined as interactions in which a participant interacts with other participants or with the teacher. In general, students often interact with their peers or teacher in a technology-assisted learning environment. Therefore, this study classifies the participant interactions of the reviewed papers into three categories: *student-teacher*, *student-student*, and *both*. For example, the participant interaction type of a teacher conducting a lecture with an electronic whiteboard was coded as *student-teacher* interaction (e.g., Lopez, 2010). Those studies in which students collaboratively played a game (e.g., Susaeta et al., 2010) or discussed together online (e.g., Wei & Chen, 2006) were coded as student-student interactions. A study including both of the abovementioned kinds of interaction was coded as “both” (e.g., Tsai, 2010).

## Inter-coder reliability

A total of six experts with doctoral degrees or professorship in the instructional technology field participated as the coders. To achieve satisfactory inter-coder reliability, the coders discussed the contents and details of the above coding scheme to reach consensus prior to the official coding. Afterwards 30 articles were randomly extracted from the *Journal of Computer Assisted Learning*, and all the coders coded the articles based on the aforementioned scheme, and multi-coder reliability tests were conducted based on the results of all the coding items. The multi-coder Fleiss Kappa reliability of all items was between 0.27 and 0.44, thus reaching the fair agreement level ( $\kappa=0.21\sim0.4$ ) (Landis & Koch, 1977), indicating that there was an acceptable inter-coder reliability among all six coders. Every paper was then assigned to be coded by an individual coder. Afterwards, all 322 articles were allocated to the six coders for coding.

## Results and discussion

### The descriptive data for the content analysis results

Table 1 shows the descriptive data for the content analysis results regarding the selected 322 papers. Among these 322 intervention studies, 104 were from 2005 to 2007 and 218 were from 2008 to 2010. The number for the last three years is double that for the first three years, indicating that an increasing number of intervention studies regarding technology-assisted instruction were conducted in the later period. However, it should also be noted that, among the 4,093 papers published in these five journals from 2005 to 2010, only a total of 322 were recognized as being empirical intervention studies on technology-assisted instruction. It seems that still relatively little research on educational technology has addressed the actual effects of specific instructional interventions on student learning.

Table 1. Descriptive data for the content analysis results

		Total n (%)	2005-2007 n (%)	2008-2010 n (%)	Chi-Square <sup>a</sup>
Sample group	Elementary school	69 (21.4)	24(23.1)	45 (20.6)	0.895 (n.s.)
	Junior and Senior high school	66 (20.5)	23(22.1)	43 (19.7)	
	Higher education	164 (50.9)	49(47.1)	115 (52.8)	
	Adults	23 (7.1)	8(7.7)	15 (6.9)	
Subject domains	Science	98 (30.4)	34 (32.7)	64 (29.4)	2.125 (n.s.)
	Mathematics	31 (9.6)	11 (10.6)	20 (9.2)	
	Art & Languages	29 (9.0)	8 (7.7)	21 (9.6)	
	Social studies	67 (20.8)	24 (23.1)	43 (19.7)	
	Engineering & Computer	74 (23.0)	21 (20.2)	53 (24.3)	
	Others	10 (3.1)	3 (2.9)	7 (3.2)	
	Unclear	13 (4.0)	3 (2.9)	10 (4.6)	
Research focus (foci)	Achievement (Ac)	156 (48.4)	45 (43.3)	111 (50.9)	4.052 (n.s.)
	Learning process (Pr)	50( 15.5)	20 (19.2)	30 (13.8)	
	Affective domain (Af)	22 (6.8)	7 (6.7)	15 (6.9)	
	Ac, Pr	28 (8.7)	11 (10.6)	17 (7.8)	

	Ac, Pr, Af	9 (2.8)	3 (2.9)	6 (2.8)	
	Ac, Af	46 (14.3)	13 (12.5)	33 (15.1)	
	Pr, Af	11 (3.4)	5 (4.8)	6 (2.8)	
Technology adoption	Technology for general purpose	113 (35.1)	41 (39.4)	72 (33.0)	1.264
	Technology for specific instructional purpose	209 (64.9)	63 (60.6)	146 (67.0)	(n.s.)

<sup>a</sup> The comparison between 2005-2007 and 2008-2010; n.s: non-significant

### Sample groups in the reviewed studies

The analysis of sample group is shown in Table 1. Regarding the distribution percentages of the sample groups analyzed from 2005 to 2010, higher education (50.9%) ranked highest, followed by high school (20.5%), elementary school (21.4%), and adult education (7.1%). This ranking sequence is the same for both the 2005-2007 and 2008-2010 sub-groups. The results above may indicate that technology-assisted instruction studies have tended to be conducted more often in higher education contexts. One possible interpretation is that most of the researchers in this field are academic scholars, and it may be easier for them to recruit their own students as participants in their experiments; in contrast, adults are less accessible to academic researchers, and thus may have limited opportunities to participate in such kinds of studies. Similar perspectives have also been proposed by Hsu et al. (2012). However, it should be noted that the finding above was derived from the five key journals reviewed in this study, which may be insufficient to generalize the findings to the entire field of educational technology research. In the last decade, the importance of fostering lifelong learning, including continuing education or employees' professional development, has been recognized by educators worldwide (Lüftenegger et al., 2012). Recently, some researchers in educational technology have also advocated the use of educational technology for promoting lifelong learning (e.g., Min, 2008). However, this study reveals that researchers have conducted relatively little research on the use of technology-assisted instruction in adult learning. It is suggested that more research on the use of technology-assisted instruction in continuing education is urgently needed.

### Subject domains involved in the reviewed studies

The most studied subject domain is science (30.4%), followed by engineering & computers (23.0%), social studies (20.8%), mathematics (9.6%), and art & language (9.0%) for papers published in the period of 2005 to 2010 as shown in Table 1. Though the ranking sequence is slightly different between the periods of 2005-2007 and 2008-2010, the ratio of each subject domain did not change significantly according to a chi-square test.

The results above show that "science" and "engineering and computers" are the two major subject domains that were studied in those papers. The popularity of science as a subject domain may be attributed to its distinct nature. On the one hand, many scientific concepts seem to be more abstract and thus are difficult to elaborate just in words. This difficulty could be overcome by the adoption of technology, such as using simulation and animation to explicitly demonstrate related scientific models and/or phenomena. On the other hand, learning science involves the learning of scientific methods or inquiry processes (e.g., Kong, Yeung, & Wu, 2009). With the aid of technology, science educators can create a better learning environment for facilitating scientific experiments which allow students to practice scientific skills. The widespread adoption of the engineering and computer domains may be accounted for by the availability of participants to the researchers. Since many academic researchers in instructional technology are computer scientists, it would be much easier for them to examine the effectiveness of the instructional technology they have developed by inviting their own students to participate directly.

### Research Foci of the reviewed studies

Regarding the research foci of the reviewed papers, the majority (74%, including the categories of "Ac," "Ac, Pr," "Ac, Af," and "Ac, Pr, Af") focus on learning achievement as a part of or as the only research outcome. However, it should be noted that relatively fewer studies investigated students' learning processes (30%, combining the categories of "Pr," "Ac, Pr," "Pr, Af," and "Ac, Pr, Af") or affective outcomes (27%, totaling the categories of "Af,"

“Ac, Af,” “Pr, Af,” and “Ac, Pr, Af”). Besides, studies covering all three categories were very rare (3%, “Ac, Pr, Af”). The Chi-square analyses on the research foci for the two periods (i.e., 2005-2007 and 2008-2010) indicate no significant differences in any of the categories. Researchers have suggested the importance of using high density observation to understand how changes would occur in a learning process (Shih, Feng, & Tsai, 2008). The findings above suggest that future research should focus not only on participants’ learning achievements but also on their learning processes and affective aspects, as advocated by Kuiper, Volman and Terwel (2005).

### **Technology adoption in the reviewed studies**

Table 1 shows that 64.9% of the studies adopted technologies for a specific instructional purpose, while only 35.1% used technologies for general purposes (such as Power Point). Technology for specific instructional purposes is more commonly adopted than technology for general purposes for both of the two groups (2005-2007 and 2008-2010).

The higher adoption rate of technology for specific instructional purposes in the reviewed studies may be due to some practical concerns. These technologies were developed for specific learning environments and activities, and thus one might expect them to be more practically useful and to enhance learning and instruction. In addition, since most technologies for specific instructional purposes were developed to be capable of automatically recording students’ learning log files, it is easier for these educational researchers to collect essential and meaningful data. When common technologies or software for general purpose were used, researchers may not have full access to these records for advanced analyses. Moreover, some educators might have been dissatisfied with the limited functions of the contemporary technology for general purposes, and thus attempted to design some pedagogical-oriented technologies, systems or software for instructional purposes to fulfill their specific instructional needs.

### **Interpersonal interaction types in the reviewed studies**

Among the 322 reviewed papers, 199 studies involved technology-assisted-interpersonal interaction, while the remaining 123 only involved human-technology interaction. In this study, the technology-assisted interactive types were divided into three categories: face-to-face interaction (FTF), technology-mediated interaction (TMI), and blended interaction. Among the 199 studies which had technology-assisted-interpersonal interaction, technology-mediated interactions (58.8%) are more popular than face-to-face interactions (20.1%) and blended interactions (21.1%) as shown in Table 2. The results reveal that technology-mediated interactions are commonly utilized in technology-assisted instruction. The popularity of technology-mediated interactions in technology-assisted instruction may be attributed to the rapid advancement of technologies in recent years. For interpersonal interactions, new technologies provide abundant tools and media. For example, instant messages, conferencing meetings, and interactive whiteboards can provide rapid and convenient channels for synchronized interaction; emails, forums, and short message systems can provide synchronous or asynchronous interactions. It is not easy to improve most of the interpersonal interactions mentioned above in traditional learning contexts.

In the Web 2.0 era, various novel information technologies or tools have been widely used in technology-assisted instruction to promote interpersonal interactions. However, it should be noted that, among the reviewed 322 papers in the current study, 123 (38%) did not involve the use of technology to promote interpersonal interactions. That is, only human-technology interactions were involved in these studies. Moreover, a slightly growing trend in face-to-face interaction from 2005 to 2010 was also observed (from 14.5% in 2005-2007 to 23.1% in 2008-2010). It seems that face-to-face interaction can provide more information, such as facial expressions, body language, and emotions. Thus, face-to-face interaction still has its irreplaceable value, a perspective revealed in social presence theory (Short, Williams, & Christie, 1976). In other words, face-to-face interactions should also be highlighted in technology-supported learning environments.

### **Participant interactions in the reviewed studies**

This study further explored participant interactions in the reviewed papers. The distribution of participant interactions is shown in Table 2. The results show that the focus was on student-student interactions (59.8%), followed by both interactions (29.6%), while very few studies (10.6%) investigated student-teacher interactions only.

The results reveal that most studies focused on student-student interactions. This may be due to the fact that the change of pedagogical paradigm has influenced teachers' application of new technologies in their instruction. In recent years, the pedagogical paradigm has shifted from teacher-centered to student-centered or constructivist-oriented teaching. In a so-called traditional classroom, the interactions mostly occur between the teacher and students. With the aid of technologies, however, students now have more opportunities to interact with each other. Educational researchers have proposed that technology can foster more student-centered learning (Ringstaff & Kelley, 2002) which is echoed by the findings in this study. Additionally, technologies could help to extend the period of interaction from in-class to out of class settings; thus the places of interaction are no longer limited to the classroom. This flexibility allows students to have more opportunities to interact with both peers and teachers. Nevertheless, in constructivist-oriented learning environments, the teacher may still play an essential role in the learning process (Chrenka, 2001). Despite this important role, based on the reviewed papers in this study, teacher-student interactions were relatively less emphasized. Further research may address teacher-student interactions under the constructivist pedagogical paradigm.

*Table 2.* Interpersonal interaction types and participant interactions revealed in the reviewed intervention studies on technology-assisted instruction from 2005-2010

		Total n (%)	2005-2007 n (%)	2008-2010 n (%)	Chi-Square <sup>a</sup>
Interpersonal interactive type	FTF (n=40)	40 (20.1)	10 (14.5)	30 (23.1)	2.901
	TMI (n=117)	117 (58.8)	41 (59.4)	76 (58.5)	(n.s.)
	Blended (n=42)	42 (21.1)	18 (26.1)	24 (18.5)	
Participant interaction	Student-student (n=119)	119 (59.8)	43 (62.3)	76 (58.5)	1.228
	Student-teacher (n=21)	21 (10.6)	5 (7.2)	16 (12.3)	(n.s.)
	Both (n=59)	59 (29.6)	21 (30.4)	38 (29.2)	

<sup>a</sup> The comparison between 2005-2007 and 2008-2010

TMI: Technology-mediated Interaction

n.s: non-significant

## Cross analysis results

In recent years, the development of new technologies has generally aimed to support user customization (Linn, 2003). In other words, when adopting technology-assisted learning, the features of subject domains and characteristics of learners should be taken into account. Moreover, the use of technologies may have changed the ways interpersonal interactions occur. Therefore, this study further conducted cross analyses between technology adoption and sample group, technology adoption and subject domain, as well as interpersonal interaction type and participant interaction.

### *Technology adoption vs. sample group*

The associations between technology adoption and sample group are shown in Table 3. In general, technology adoption is significantly associated with sample group for the period of 2005-2010 ( $\chi^2 = 14.9, p < .05$ ). In addition, among the studies that used adults as their samples, technology for general purposes is adopted significantly more often than technology for a specific instructional purpose ( $AR = 3.1$ ). Although the association between technology adoption and sample group is not significant for the period of 2005-2007, an association is found in the period of 2008-2010 ( $\chi^2 = 11.9, p < .05$ ), as shown in Table 4. Specifically, among the studies that used adults as their samples, technologies were more often adopted for general purposes ( $AR = 2.3$ ) rather than for a specific instructional purpose in the period of 2008-2010. In contrast, among the studies which used high school students as their samples, technology for a specific instructional purpose ( $AR = 2.2$ ) is adopted significantly more often than technology for general purposes. It seems that high school students have to learn domain-specific learning tasks; as a result, technologies for specific instructional purposes were adopted to assist their learning.

The reason why studies with adults as participants tended to use technology for general purposes is likely to be that adults are more familiar with such technologies. Technologies for general purposes can meet adults' ordinary usage needs, and require less technical competence and assistance.



*Table 3.* Frequencies and adjusted residual between technology adoption and sample group from Chi-square analysis (2005-2010)

	Elementary School Students	High School Students	Higher Education Students	Adults
	(n, AR)	(n, AR)	(n, AR)	(n, AR)
Technology for general purpose (n=113)	(18, -1.8)	(17, -1.8)	(63, 1.3)	(15, 3.1)
Technology for specific instructional purpose (n=209)	(51, 1.8)	(49, 1.8)	(101, -1.3)	(8, -3.1)
Pearson's Chi-Square	14.94*			

AR: Adjusted residual values (AR with absolute values larger than 1.96 are significant.)

\* $p < 0.05$

*Table 4.* Frequencies and adjusted residual between technology adoption and sample group from Chi-square analysis (2008-2010)

	Elementary School Students	High School Students	Higher Education Students	Adults
	(n, AR)	(n, AR)	(n, AR)	(n, AR)
Technology for general purpose (n=113)	(11, -1.4)	(8, -2.2)	(44, 1.7)	(9, 2.3)
Technology for specific instructional purpose (n=209)	(34, 1.4)	(35, 2.2)	(71, 1.7)	(6, -2.3)
Pearson's Chi-Square	11.90*			

AR: Adjusted residual values (AR with absolute values larger than 1.96 are significant.)

\* $p < 0.05$

### Technology adoption vs. subject domain

The association between technology adoption and subject domain is shown in Table 5. There is a significant association between technology adoption and subject domain for the period of 2005 to 2010 ( $\chi^2 = 14.79$ ,  $p < .05$ ). For the studies whose subject domain was science, more technologies for a specific instructional purpose (AR = 3.1) were adopted than for general purposes. In contrast, for the studies that focused on the “engineering & computer” domain, more technologies for general purposes were adopted (AR = 2.5) than technology for a specific instructional purpose. It should be noted that, for the studies that focused on the “engineering & computer” domain in the period of 2005-2010, 35 used technology for general purposes while 39 used technology for a specific instructional purpose. However, the total number of studies using technology for general purposes was 113, while 209 used technology for a specific instructional purpose in the period of 2005-2010. Thus, the percentage of studies focusing on the “engineering & computer” domain among those using technology for general purposes is significantly higher than the average. However, the associations are not significant in the two sub groups of 2005-2007 and 2008-2010.

The results indicate an association between technology adoption and subject domain. Regarding science, as mentioned previously, since most scientific concepts are abstract in nature and the scientific methods are complex, specially-designed technology for a specific instructional purpose can better support students to learn related scientific concepts and methods. For example, to enhance students' science learning and application, Fund (2007) used a computerized environment that creates a microworld with problem-solving scaffolding supports to help students conduct simulated-based experiments. In contrast, since most engineering or computer courses are designed for helping students learn basic computer programs or algorithms, technology for general purposes is more likely to be adopted as a basic “tool” for assisting learning.

Table 5. Frequencies and adjusted residual between technology adoption and subject domain from Chi-square analysis (2005-2010)

	Subject domain						
	Science	Math	Arts & Languages	Social Study	Engineering & Computer	Others	Non-Specified
	(n, AR)	(n, AR)	(n, AR)	(n, AR)	(n, AR)	(n, AR)	(n, AR)
Technology for general purpose (n=113)	(22,-3.1)	(9,- 0.7)	(9, -0.5)	(28, 1.3)	(35, 2.5)	(3, 1.0)	(5, 0.3)
Technology for specific instructional purpose (n=209)	(76, 3.1)	(22, 0.7)	(20, 0.5)	(39, -1.3)	(39, -2.5)	(5, -1.0)	(8, -0.3)
Pearson's Chi-Square	14.79*						

AR: Adjusted residual values (AR with absolute values larger than 1.96 are significant.)

\* $p < 0.05$

### Interpersonal interaction type vs. participant interaction

As shown in Table 2, a total of 199 studies have employed technology-assisted interpersonal interaction. To further understand the association between interaction type and participant interaction in the studies, a Pearson Chi-square test was performed. There is a significant association between interactive types and participant interactions for the studies during 2005 to 2010 (Chi-square = 10.57,  $p < 0.05$ ) as shown in Table 6. In particular, the studies with participant interactions including both student-student and student-teacher had a strong tendency to adopt the blended mode (F2F and TMI) of interpersonal interactions.

To enhance the effects of instruction, various learning activities of technology-assisted interpersonal interactions were designed. For example, in traditional learning environments, the opportunities for peer (i.e., student-student, SS) interaction are relatively limited. Thus, studies in this line may make additional efforts to design various learning activities with technology-mediated interaction (e.g., Gijlers, Saab, van Joolingen, de Jong, & van Hout-Wolters, 2009; Gilbert & Dabbagh, 2005; Schellens & Valcke, 2006). Moreover, the lecture is still a dominant form for student-teacher (ST) interpersonal interaction in a traditional classroom. Thus, extra efforts have been made to develop learning activities with blended interpersonal interaction that may enhance the interaction between students and teachers (e.g., Lim, Reiser, & Olina, 2009; Ozmen, 2008). In addition, the associations between interpersonal interaction types and participant interactions are not significant in the two sub groups of 2005-2007 and 2008-2010.

Table 6. Frequencies and adjusted residual between the interpersonal interaction types and the participant interactions of the studies from Chi-square analysis (2005-2010)

	Student versus Student (SS)	Student versus Teacher(ST)	Both (SS and ST)
	(n, AR)	(n, AR)	(n, AR)
Face To Face (FTF) (n=40)	(23, -0.3)	(7, 1.6)	(10, -0.7)
Technology-Mediated Interaction (TMI) (n=117)	(76, 1.8)	(12, -0.2)	(29, -1.8)
Blended (FTF and TMI) (n=42)	(20, -1.8)	(2, -1.4)	(20, 2.9)
Pearson's Chi-Square	10.57*		

AR: Adjusted residual values (AR with absolute values larger than 1.96 are significant.)

\* $p < 0.05$

### Conclusions and directions for further research

Pedagogical considerations are crucial for the use of technology in education (Leijen et al., 2008). By reviewing papers published in five important SSCI journals from 2005 to 2010, this study aimed to provide insights into intervention studies on technology-assisted instruction. It should be acknowledged that while the five journals selected are major journals in the field of educational technology, there are many other studies pertaining to

education technology published elsewhere. Therefore, the interpretation of the results derived from this study should be made with care. Besides, more journals for educational technology are suggested to be used as the literature source in further research. Or, via certain keyword searching in some electronic databases (e.g., SSCI or Google Scholar), the researchers can collect more articles in various journals or conferences as the literature source for further research. Thus, a more complete picture regarding intervention studies on technology-assisted instruction can be provided.

The major research trends derived from the results of this study are summarized and concluded below. In this study, a remarkable increase in the number of empirical studies on technology-assisted instruction from 2005 to 2010 was found. However, it should also be noted that, among the 4,093 papers published in these five journals from 2005 to 2010, only a total of 322 were identified as intervention studies regarding technology-assisted instruction. This study suggests that more attention should be paid to the role of interventions in technology-assisted instruction in future empirical research. Moreover, this study also found that very few studies have simultaneously addressed achievement, learning process, and affective outcomes. This suggests that further research on technology-assisted instruction may be conducted with various samples, different subject domains, or multiple research foci. Moreover, digital literacy outcomes and metacognitive knowledge have received an increasing amount of attention in recent years (e.g., Jimoyiannis & Gravani, 2011; Topcu & Ubuz, 2008). Future research can also include these outcomes when analyzing the research foci among the reviewed studies.

Regarding technology adoption, this study reveals that, in both the 2005-2007 and 2008-2010 periods, technologies for specific instructional purposes were more frequently adopted than those for general purposes. This study also analyzed interpersonal interactions within empirical research on technology-assisted instruction. It was revealed that technology-mediated interpersonal interactions were relatively commonly utilized in these intervention studies and that these interactions mostly focused on those among peers. Finally, the cross analyses in this study showed that technology for general purposes were frequently used in the reviewed studies with adults as participants.

In sum, the findings derived from the current study provide important insights into the effects and research tendencies of these relatively more informative studies. One may be interested in the specific instructional strategies used in these intervention studies on technology-assisted instruction reviewed in this study. However, when conducting content analyses, we found that the strategies used in some studies were not clearly described; besides, a specific instructional strategy might be introduced using a variety of terminologies in different studies. Therefore, it is very difficult to categorize the strategies (such as scaffolding, inquiry) used in these intervention studies on technology-assisted instruction during the period of 2005 to 2010. This study suggests that further studies regarding technology-assisted instruction should be carried out with specific technology-assisted instructional activities, and that more in-depth data from participants should be collected. Thus, the actual influences of using technology in instructional practices can be further explored.

## Acknowledgements

Funding for this research work is supported by the National Science Council, Taiwan, under grant numbers NSC 101-2911-I-011-510, NSC 100-2631-S-011-001, 101-2631-S-011-001 and NSC101-2911-I-110-508.

## References

- Alonso, F., Manrique, D., & Viñes, J. M. (2009). A moderate constructivist e-learning instructional model evaluated on computer specialists. *Computers & Education*, 53(1), 57-65.
- Balram, S., & Dragicevic, S. (2008). Collaborative spaces for GIS-based multimedia cartography in blended environments. *Computers & Education*, 50(1), 371-385.
- Bernard, R. M., Abrami, P. C., Borokhovski, E., Wade, C. A., Tamim, R. M., Surkes, M. A., & Bethel, E. C. (2009). A meta-analysis of three types of interaction treatments in distance education. *Review of Educational Research*, 79(3), 1243-1289.
- Bernard, R. M., Abrami, P. C., Lou, Y., & Borokhovski, E. (2004). A methodological morass? How we can improve the quality of quantitative research in distance education. *Distance Education*, 25(2), 176-198.

- Cho, K., & Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers & Education*, 48(3), 409-426.
- Chrenka, L. (2001). Misconstruing constructivism. *Phi Delta Kappan*, 82(9), 694-695
- Dillon, A., & Gabbard, R. (1998). Hypermedia as an educational technology: A review of the quantitative research literature on learner comprehension, control, and style. *Review of Educational Research*, 68(3), 322-349.
- Fabos, B., & Young, M. D. (1999). Telecommunication in the classroom: Rhetoric versus reality. *Review of Educational Research*, 69(3), 217-259.
- Fessakis, G., Tatsis, K., & Dimitracopoulou, A. (2008). Supporting "Learning by design" activities using group blogs. *Educational Technology & Society*, 11(4), 199-212.
- Fund, Z. (2007). The effects of scaffolded computerized science problem-solving on achievement outcomes: a comparative study of support programs. *Journal of Computer Assisted Learning*, 23(5), 410-424.
- Gilbert, P. K., & Dabbagh, N. (2005). How to structure online discussions for meaningful discourse: A case study. *British Journal of Educational Technology*, 36(1), 5-18.
- Gijlers, H., Saab, N., van Joolingen, W. R., de Jong, T., & van Hout-Wolters, B. H. A. M. (2009). Interaction between tool and talk: How instruction and tools support consensus building in collaborative inquiry-learning environments. *Journal of Computer Assisted Learning*, 25, 252-267.
- Grimshaw, S., Dungworth, N., McKnight, C., & Morris, A. (2007). Electronic books: children's reading and comprehension. *British Journal of Educational Technology*, 38(4), 583-599.
- Hewitt, J. (2003). How habitual online practices affect the development of asynchronous discussion threads. *Journal of Educational Computing Research*, 28(1), 31-45.
- Hewitt, J. (2005). Toward an understanding of how threads die in asynchronous computer conference. *The Journal of Learning Science*, 14(4), 567-589.
- Hoskins, S. L., & van Hooff, J. C. (2005). Motivation and ability: Which students use online learning and what influence does it have on their achievement? *British Journal of Educational Technology*, 36, 177-192.
- Hou, H. T., & Wu, S. Y. (2011). Analyzing the social knowledge construction behavioral patterns of an online synchronous collaborative discussion instructional activity using an instant messaging tool: A case study. *Computers & Education*, 57(2), 1459-1469.
- Hsu, Y. C., Ho, H. N. J., Tsai, C. C., Hwang, G. J., Chu, H. C., Wang, C. Y., & Chen, N. S. (2012). Research Trends in Technology-based Learning from 2000 to 2009: a content analysis of publications in selected journals. *Educational Technology & Society*, 15(2), 354-370.
- Jimoyiannis, A., & Gravani, M. (2011). Exploring adult digital literacy using learners' and educators' perceptions and experiences: The case of the second chance schools in Greece. *Educational Technology & Society*, 14(1), 217-227.
- Kong, S. C., Yeung, Y. Y., & Wub, X. Q. (2009). An experience of teaching for learning by observation: Remote-controlled experiments on electrical circuits. *Computers & Education*, 52(3), 702-717.
- Kuiper, E., Volman, M., & Terwel, J. (2005). "The web as an information resource in K-12 education: strategies for supporting students in searching and processing information." *Review of Educational Research*, 75(3), 285-328.
- Landis, J. R., & Koch, G. G. (1977). "The measurement of observer agreement for categorical data." *Biometrics*, 33(1), 159-174.
- Lim, J., Reiser, R. A., & Olina, Z. (2009). The effects of part-task and whole-task instructional approaches on acquisition and transfer of a complex cognitive skill. *Educational Technology Research and Development*, 57(1), 61-77.
- Linn, M. C. (2003). Technology and science education: Start point, research program, and trends. *International Journal of Science Education*, 25, 727 – 758.
- Lawless, K. A., & Pellegrino, J. W. (2007). Professional development in integrating technology into teaching and learning: Knowns, unknowns, and ways to pursue better questions and answers. *Review of Educational Research*, 77(4), 575-614.
- Lee, S. W. Y., & Tsai, C.-C. (2011). Identifying patterns of collaborative knowledge exploration in online asynchronous discussions. *Instructional Science*, 39(3), 321-347.
- Lee, S. W. Y., Tsai, C.-C., Wu, Y.-T., Tsai, M.-J., Liu, T. C., Hwang, F.-K., ... Chang, C.-Y. (2011). Internet-based science learning: A review of journal publications. *International Journal of Science Education*, 33(14), 1893-1925.
- Leijen, Ä., Admiraal, W., Wildschut, L., & Simons, P. R. J. (2008). Pedagogy before technology: What should an ICT intervention facilitate in practical dance classes? *Teaching in Higher Education*, 13(2), 219-231.

- Liu, C.-C., & Tsai, C.-C. (2008). An analysis of peer interaction patterns as discoursed by on-line small group problem-solving activity. *Computers & Education*, 50(3), 627-639.
- Looi, C. K., Chen, W. L., & Ng, F. K. (2010). Collaborative activities enabled by GroupScribbles (GS): An exploratory study of learning effectiveness. *Computers & Education*, 54(1), 14-26.
- Lopez, O. S. (2010). The digital learning classroom: Improving English language learners' academic success in mathematics and reading using interactive whiteboard technology. *Computers & Education*, 54(4), 901-915.
- Lüftenegger, M., Barbara, S., van de Schoot, R., Wagner, P., Finsterwald, M., & Spiel, C. (2012). Lifelong learning as a goal: Do autonomy and self-regulation in school result in well prepared pupils? *Learning and Instruction*, 22(1), 27-36.
- Min, Y. (2008). Rethinking lifelong learning through online distance learning in Chinese educational policies, practices and research. *British Journal of Educational Technology*, 39, 583-597.
- Monteserin, A., Schiaffino, S. N., & Amandi, A. (2010). Assisting students with argumentation plans when solving problems in CSCL. *Computers & Education*, 54(4), 416-426.
- Ozmen, H. (2008). The influence of computer-assisted instruction on students' conceptual understanding of chemical bonding and attitude toward chemistry: A case for Turkey. *Computers & Education*, 51(1), 423-438.
- Ringstaff, C., & Kelley, L. (2002). *The learning return on our educational technology investment: A review of findings from research*. Washington, DC: Office of Educational Research and Improvement.
- Schellens, T., & Valcke, M. (2006). Fostering knowledge construction in university students through asynchronous discussion groups. *Computers & Education*, 46(4), 349-370.
- Shih, M, Feng, R., & Tsai, C.-C. (2008). Research and trends in the field of e-learning from 2001 to 2005: A content analysis of cognitive studies in selected journals. *Computers & Education*, 51(2), 955-967.
- Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. London, England: John Wiley.
- Straub, E. T. (2009). Understanding technology adoption: Theory and future direction for informal learning. *Review of Educational Research*, 79(2), 625-649.
- Susskind, J. E. (2008). Limits of PowerPoint's power: Enhancing students' self-efficacy and attitudes but not their behavior. *Computers & Education*, 50(4), 1228-1239.
- Susaeta, H., Jimenez, F., Nussbaum, M., Gajardo, I., Andreu, J. J., & Villalta, M. (2010). From MMORPG to a classroom multiplayer presential role playing game. *Educational Technology & Society*, 13(3), 257-269.
- Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C., & Schmid, R. F. (2011). What forty years of research says about the impact of technology on learning: A second-order meta-analysis and validation study. *Review of Educational Research*, 81(1), 4-28.
- Topcu, A., & Ubuz, B. (2008). The effects of meta-cognitive knowledge on the preservice teachers' participation in the asynchronous online forum. *Educational Technology & Society*, 11(3), 1-12.
- Truman, S. M., & Truman, P. J. (2006). An investigation of the situated learnability effects of single- and dual-modal systems in education: a report of music-oriented learning environment and science computer-assisted teaching studies. *British Journal of Educational Technology*, 37(1), 131-142.
- Tsai, C. W. (2010). Do students need teacher's initiation in online collaborative learning? *Computers & Education*, 54(4), 1137-1144.
- van Rooij, S. W. (2009). Adopting open-source software applications in US higher education: A cross-disciplinary review of the literature. *Review of Educational Research*, 79(2), 682-701.
- Wei, F. H., & Chen, G. D. (2006). Collaborative mentor support in a learning context using a ubiquitous discussion forum to facilitate knowledge sharing for lifelong learning. *British Journal of Educational Technology*, 37(6), 917-935.
- Wever, B. D., Schellens, T., Valcke, M., & Keer, H. V. (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Computers & Education*, 46(1), 6-28.
- Zydney, J. M. (2010). The effect of multiple scaffolding tools on students' understanding, consideration of different perspectives, and misconceptions of a complex problem. *Computers & Education*, 54(2), 360-370.

## Development of a Virtual Second Life Curriculum Using Currere Model

Chih-Feng Chien<sup>1\*</sup>, Trina Davis<sup>2</sup>, Patrick Slattery<sup>3</sup>, Wendy Keeney-Kennicutt<sup>4</sup> and Janet Hammer<sup>5</sup>

<sup>1</sup>Graduate School of Education, Chung Yuan Christian University, Taoyuan 32023, Taiwan, R.O.C. // <sup>2</sup>Department of Teaching, Learning and Culture, Texas A&M University, TX 77843, USA // <sup>3</sup>Department of Teaching, Learning and Culture, Texas A&M University, TX 77843 USA // <sup>4</sup>Department of Chemistry, Texas A&M University, TX 77843, USA // <sup>5</sup>Department of Teaching, Learning, and Culture, Texas A&M University, TX 77843 USA //

iorifeng@gmail.com // trinadavis@tamu.edu // patslat@aol.com // kennicutt@chem.tamu.edu // jhammer@tamu.edu

\*Corresponding author

(Submitted March 18, 2012; Revised July 05, 2012; Accepted October 17, 2012)

### ABSTRACT

This is a study integrating Second Life (SL) and the *currere* approach to develop a virtual curriculum demonstration. The overarching purposes of this study were to understand the perceptions, self-reflection, self-understanding and educational growth of graduate students in education toward teaching and learning in a virtual interdisciplinary curriculum. Data was collected in two education graduate courses with a total of 31 participants in 2011 at a central Texas public university. The results revealed how SL exhibitions, based on the four-step *currere* approach, benefitted the participants. Using the exhibitions' target knowledge, individuals were able to develop a self-understanding, which propelled them toward self-mobilization and educational reconstruction.

### Keywords

Second Life, *Currere*, Self-reflection, Self-understanding, Educational reconstruction

### Introduction

The idea of a virtual-world exhibition was inspired by a physical Vietnam War exhibit. The exhibition included collages of pictures and documentary videos that were designed by students in a doctoral-program curriculum theory course at a public university. The exhibition was next taken to a Curriculum and Pedagogy Conference. After the experiences with two exhibitions, the researcher extended this curriculum to a virtual world (VW) for increased multimedia and substantial curriculum.

The study utilizes two Second Life (SL) installations to deliver a meaningful curriculum on the basis of William Pinar's (2004) *currere* approach—a curriculum technique used to reconstruct social, intellectual, and physical structures. The study is significant in three ways:

- It contributes to research into SL and its application to the *currere* approach to develop curriculum.
- It demonstrates VW curriculum development.
- It contributes to the understanding of education major graduate students' perceptions of various perspectives.

The VW environment has been used in a variety of general curricula, such as science (Bainbridge, 2007), mathematics (Roussou, 2009), architecture (Clark & Maher, 2001), etc. However, little to no VW research has been conducted using the *currere* approach, the curriculum theory Pinar describes as “the interdisciplinary study of educational experience” (Pinar, 2004, p. 2). The *currere* approach is a curriculum technique used to reconstruct social, intellectual, and physical structures (Grumet, 1976). This is an autobiographic process used to develop a curriculum sufficient enough to engage students in self and mutual conversation, what Pinar (2004) and Grumet (1976) called complicated conversation and autobiographic narrative. This approach, if examined regressively, progressively, analytically and synthetically, can engage students in complicated conversation and autobiographic narratives. It also has the potential to elicit self-understanding and self-reflection thereby helping teachers and students to comprehend the true meaning of a curriculum. Research on virtual curriculum development in the *currere* approach has never been done. Hence, this research which used the *currere* approach to engage education graduate students in reflections on their perceptions, understanding, and professional development in virtual exhibitions, contributes to the research of virtual curriculum development.

Secondly, one of the purposes for studying virtual curriculum development is to demonstrate its theoretical and practical aspects. Some participants have had experiences teaching in primary and secondary schools. By using multimedia technology in VWs to deliver a curriculum, some participants learned to enlarge their knowledge of implementing curriculum in a creative way; some understood education can be done in synthetic, analytic, and dialogic ways. This demonstration eventually benefited some of the participants in their current or future teaching endeavors.

Finally, the participants experienced two past-to-future exhibitions, in which they were expected to extract their inner voice. The inner voice here reflects from the “historical and natural world” (Pinar, 2004, p. 37). A part of the study’s contribution is to identify several perspectives for understanding how to use VWs to teach appropriately. By using SL as a visual-based and auditory-based simulated environment, this study adopted William Pinar’s *currere* approach to examine students’ self-understanding and self-reflection in order to achieve the true meaning of curriculum.

### Research questions

The study specifically seeks to answer the following research questions:

- What are the participants’ perceptions about teaching and learning in SL curricula?
- What are the participants’ self-reflections and self-understanding in the virtual curricular demonstration?
- How do the participants describe their educational growth and development at the end of their SL experiences?

### Literature review

#### Currere

The theoretical framework for developing this study is based mainly on the *currere* approach. According to Pinar (2004), *currere* is the Latin infinitive form which means *to run the curriculum*, or in gerund form meaning *the running of the curriculum*. Sameshima and Irwin (2008) especially address that “curriculum is static, while *currere* is dynamic” (p. 7). Irwin (2006) states “*currere* is the active form of curriculum; a *currere* emphasizes acts of inquiry over a course of action” (p. 75). *Currere* is the method used to investigate the relationship between academic knowledge, life history, self-understanding, and social reconstruction (Pinar, 2004).

Pinar (2004) suggests four essential stages of the *currere* approach: (1) the regressive stage: to return to the past, enlarge one’s memories, and transform them to present; (2) the progressive stage: to look toward what is not yet the case and what is not yet present; (3) the analytical stage: to examine “one’s distanciation from past and future functions in order to create a subjective space of freedom in the present,” and (4) the synthetic stage: to listen carefully to “one’s own inner voice in the historical and natural world” to achieve the state of ultimate self-understanding, self-reflection, and self-transformation (pp. 35-37).

The initial *currere* research was executed by Pinar and Grumet in 1975. Working with two assistants and eleven student teachers at the University of Rochester, they first utilized *currere* to design a teachers’ training seminar. In the three-week series of the teacher-training seminar, Pinar and Grumet used *currere* to examine student teachers’ responses and educational experience. The entire course was structured as a metaphor of everyone’s own experience—sharing mutual educational experiences in order to awake individual experience. The purpose was to make student teachers discover their own experiences so they could bring them into the high school curriculum (Grumet, 1976). After Pinar, Reynolds, Slattery and Taubman published the book *Understanding Curriculum* in 1995; the *currere* method in curriculum study and research has been widely recognized and popularized. There are more than 150 peer-reviewed journals, 30 books published and many doctoral dissertations related to the *currere* approach.

## **Virtual worlds in teaching and learning**

A VW is an online computer simulation program in which users are represented by virtual surrogates. VWs comprise several features for teaching and learning in education: (1) accommodation of web-based and face-to-face classes, (2) visual learning, (3) multimedia teaching and learning, (4) benefits for disabled learners, and (5) real-world and imaginative environmental immersions.

Thompson (2008) attributes a number of factors to the trend toward more distance education: (1) the increased number of non-traditional students, (2) the increased mobility of population, and (3) the increased expectation for schools constructed to fit students' needs. The need for distance education increases the popularity of web-based online classes. However, the browser-based online class is criticized as "depersonalization and lowered bandwidth for communication" (Thompson, 2008, p. 166). Compared with traditional face-to-face classes, online classes lack human interaction and communication. Nevertheless, the downside of face-to-face classes include increasing student anxiety and lowering convenience when accessing course materials and achieving content depth (Wang & Woo, 2007). VWs, on the other hand, not only make online asynchronous discussions and assignments possible (Dickey, 2005; Kluge & Riley, 2008), but they also provide a platform for simultaneous meetings and interactions via VW users (Dickey, 2005; Dillenbourg, Schneider, & Synteta, 2002), such as avatars in SL.

VWs provide visual materials, contents, and objects (Dickey, 2005). These features broaden learning opportunities, particularly for visual learners (Annetta & Park, 2006). McComas, Pivik, and Laflamme (1998) point out that VWs can be freely customized for visual learners to understand visual information. Barab, Thomas, Dodge, Carteaux, and Tuzun (2005) indicate that VWs provide "rapid visual prototyping" for educators to understand and respond to children's preferences. Minocha and Reeves (2010) discovered that the visual realism of a learning space impacts on learners' motivation through visual learning experience and activities.

The multiple media content of VWs include visual and audio aids, video players, handouts, and outside resources connections (e.g., website, forum and blog) (Thompson, 2008). With the large-scale technologies—the skills of graphic design, audio playing, video applications, and word processing—embedded in VWs, teachers and students have the opportunities to learn and practice technologies.

VWs provide functional contents and materials for some disabled learners. McComas, Pivik, and Laflamme (1998) put forward the tactics of providing auditory information to hearing-impaired learners. In addition, learning materials can be broken down to easily understandable steps for children with learning difficulties. Furthermore, VWs also provide a barrier-free space for people with physical disabilities to participate in classroom activities like any other student (Ball & Pearce, 2009).

One of the unique advantages of teaching with VWs is their ability to represent real life scenarios, which are similar to immersive learning environments (de Freitas & Martin, 2006; Thompson, 2008). The immersive spaces simulate real-world objects, environments, and people, thus ensuring a more effective learning. Kanade and Rander (1997) said that virtual reality transcribes visual events, recovers 3D structure, and generates synthetic viewpoints. Learners can see a real-world event or object from different viewing angles at different viewing time. de Freitas, Rebollo-Mendez and Liarokapis (2010) also advocated that the immersive learning experience works better than traditional face-to-face learning. VWs' virtual reality and imaginative immersion ensure effective teaching and learning.

## **Second Life in teaching and learning**

SL, a 3D VW launched in 2003 and developed by Linden Lab, is an online, visual-based platform where multi-users interact and collaborate through mobile avatars. In education, instructors can use their avatars to interact with students using media in several ways, such as text or voice chatting with students (Boulos, Hetherington, & Wheeler, 2007), dancing with students (Gollub, 2007), and conducting presentations using a virtual slide viewer (Greenhill, 2008).

SL supports synchronous communication (de Lucia, Francese, Passero, & Tortora, 2009; Erra & Scanniello, 2009), where an instructor can hold office hours in-world (in SL) (Graves, 2008), the class can meet virtually, students can



have field trips to any world attractions and museums (Mckay, Van Schie, & Headley, 2008), and a conference can take place online (Aydogan, Aras, & Karakas, 2010). SL also supports asynchronous activities, where students engage in learning materials designed by instructors, work on virtual assignments (notecard writings, virtual presentations, creating objects, etc.), and interact with animated objects in-world. Any interactions happening in the real world can be carried out in SL.

SL supports numerous media, including images, videos, web pages, and animations (de Lucia et al., 2009; Cheal, 2007). An instructor can intentionally design virtual materials to integrate with these media so students can be fully immersed in the learning. Through synchronous or asynchronous interactions with technological media, students are engaged in inquiry-based (Vrellis, Papachristos, Bellou, Avouris, & Mikropoulos, 2010) and student-centered learning (Inman, Wright, & Hartman, 2010).

SL combines a variety of computer-mediated communication (CMC) (Ditcharoen, Naruedomkul, & Cercone, 2010). CMC is any communicative transaction taking place between two or more connected or networked computers (e.g., instant messages, emails, chat rooms, Skype, Google voice, etc.). In SL, avatars can send synchronous instant messages and also do immediate voice chat with other avatars (Boulos, Hetherington, & Wheeler, 2007). By connecting a media player, it is also possible to do video chat, though this involves more advanced technology and higher bandwidth.

## **Methodology**

This research intended to demonstrate the effect of a model curriculum which integrates with one of the most commonly used VW environments—SL, on education graduate students. Using two virtual installations—war and ecology exhibitions, participants were expected to generate their own reflections, understanding, perceptions, and opinions regarding the interdisciplinary curricula. This virtual participation included the engagement of complicated conversation and autobiographic reflection, as well as the experience of multimedia installations. This qualitative investigation sought to discover and explore how a SL curriculum reflected the participants' perceptions and practices on teaching a curriculum “across the school subjects and academic disciplines” (Pinar, 2004, p. 21).

## **Subjects**

There were two groups of participants in this study. Data from the first group were collected in the spring of 2011 and that of the second group were collected in the summer of 2011. The first group consisted of 16 graduate students who enrolled in a required course “Curriculum Development in Foundations of Education” at a large public university located in central Texas, United States. The second group consisted of 15 graduate students who enrolled in an elective course “Issues in Curriculum and Instruction” at the same university. Both courses were online courses taught by the same instructor. Five voluntary interviewees were selected from the two groups. The study was designed as an optional project for students as a substitute for a final paper. After informed about the purposes of the study, students had a week to decide on which of the two projects they would want to participate. Both participants and interviewees were recorded with the required permission signatures on IRB consent forms.

## **Process**

Data for this study were collected between February 2011 and June 2011. Students from both spring and summer courses expressed their interests and willingness to participate in this study through emails. Through a content management system, Blackboard, the researchers uploaded several handouts for training in basic SL skills. While participants were being trained in basic SL skills, the researchers held a SL help session for each group (spring and summer class) before the actual research presentations. The help sessions were held using SL and lasted for one hour, the main purpose was to discuss technical issues related to computer hardware or SL skills. Next, the researchers presented a war (see Figure 1) and an ecology exhibition (see Figure 2) on different days. Each of the presentations lasted for 1.5 hours. The presentations were recorded using Camtasia—a screen capture software.



Figure 1. War Exhibition



Figure 2. Ecological Exhibition

Following their experiences interacting with two exhibitions, participants answered a four-page open-ended survey as a reflection of their participation (see Appendix A). The survey questions focused on the four steps of the *currere* experience. Part of the survey questions were modified from the study by Brown (2007). At the end of each semester, the researchers recruited five volunteer participants for interviews. The interview questions focused more on SL perspectives (See Appendix B). Part of the interview questions were modified from the study by Jarmon, Traphagan, Mayrath and Trivedi (2009).

### Data analysis

Four forms of data—virtual observation, virtual reflective writings, individual *currere* writings, and voluntary interviews—were analyzed using interpretative phenomenological analysis (IPA). Smith, Jarman, and Osborn (1999)

stated that the goal of IPA is, instead of producing an objective statement, to explore individual personal perception, in-depth experience, and interpretation of a specific object or event.

The SL presentation recordings were written as observation reports and used as complementary resources. The interviews were transcribed verbatim from both SL and mp3 recordings. By adopting IPA to explore participants' views of the virtual curriculum development and experiences, this analysis supported the study design of *currere* for understanding graduate students' perception, reflection, and growth engaged in virtual exhibitions. Smith, et al. (1999) stated that IPA is concerned with an individual's perception, cognition, thoughts, or beliefs about a specific topic, which, in my case, was a virtual curriculum development about war and ecology exhibitions.

Modified from the IPA of Smith et al. (1999), each participant's virtual reflective writing and *currere* writing were read several times with initial thoughts and provisional codes noted on the left-hand side margin. Similar to what Eatough and Smith (2006) stated, the left-hand margin is to record anything significant and of interest. Then the specific themes transformed from initial notes and generated from multiple readings should be written in the right-hand margin (Smith & Dunworth, 2003). Interview recordings were transcribed, initially coded, and generated themes in the same way. The first step of coding the transcript was to divide up the individual extract. Each extract was a meaningful unit—one or two sentences long or a paragraph. Each extract was examined in order to generate a provisional code. Once a list of provisional codes was produced, the list was examined to see if these codes could be grouped together in meaningful ways. When the coding scheme had been produced, each extract was relabeled and given a numerical code. If a specific extract was associated with more than one category, it was given additional codes. The next step was to group the extracts based on the new coding categories. The collections of extracts were created based on the similar concepts. When the separate clusters of themes were created, the next step was to search for patterns and connections.

Table 1 is an example of virtual reflective writings analyzed using IPA.

*Table 1. The example of interpretative phenomenological analysis.*

Provisional codes	KE (Participant's name)	Themes
Notice the interlinking relations in the ecological issues.	Has anyone noticed that one of that some of the ecological problems presented here are actually interlinked to one another? For example, the receding glacier would inevitably lead to higher sea levels which would cause water levels to come further inland. As water levels move further inland, whatever is on the ground becomes a part of the water which overtime leads to contaminated water - which is another ecological issue.	Ecological relationship
		Examples
Giving two examples		
Point out other ecological issues she noticed	Also, did anyone notice that the display on contaminated water only dealt with how it affects humans? What about the animals and ecological systems that are affected?	

## Findings and discussion

The IPA drew coding schemes for virtual reflective writings, *currere* writings, and technology related data respectively, in which the first coding number presents as schemes, second number as categories, and third number as the themes.

Students' SL reflections were aggregated from the notecards written during their participation in the exhibitions. The length of each student's reflective writing (Table 2) ranged from four to eight sentences. Some people wrote more than one reflection for each exhibition while others wrote only one.

*Currere* writings were collected from all participants at the end of the two courses, ranging from seven to fifteen single-spaced pages. The data was analyzed through the same analytic process for virtual reflective writings. The *currere* writings were only read twice because of the large amount of writings. Table 3 was generated through the

four *currere* phases following the theoretical framework and questionnaire development: regression, progression, analysis, and synthesis, are considered four categories.

Table 2. Coding scheme for virtual reflective notecards

1.1 Self Disclosure	1.2 Actions	1.3 Mental Expression	1.4 Curriculum	1.5 Technology Application
1.1.1 Regressive memory 1.1.2 Regressive thinking 1.1.3 Self-reflection 1.1.4 Self-awareness 1.1.5 Prospection 1.1.6 Self-experience 1.1.7 Self-understanding	1.2.1 Importance of education 1.2.2 Taking action 1.2.3 Sense of crisis 1.2.4 Responsibilities	1.3.1 Attitudes 1.3.2 Emotions 1.3.3 Feelings 1.3.4 Imagination 1.3.5 Belief	1.4.1 Curriculum demonstration 1.4.2 Subject integration 1.4.3 Spontaneous learning	1.5.1 SL materials 1.5.2 Virtual reality

Table 3. Coding scheme for *currere* writings

2.1 Regression	2.2 Progression	2.3 Analysis	2.4 Synthesis
2.1.1 Regression memory 2.1.2 Regressive thinking 2.1.3 Regressive emotions & attitudes 2.1.4 Impact of videos 2.1.5 Regressive reflection	2.2.1 Educating 2.2.2 changing curriculum 2.2.3 Progressive imagination 2.2.4 Bringing awareness 2.2.5 Holocaust memorial	2.3.1 Continuing to educate and learn 2.3.2 self-awareness 2.3.3 Mixed feeling 2.3.4 Big ideas 2.3.5 Analytic thinking	2.4.1 Synthetic thinking 2.4.2 Self-understanding 2.4.3 Self-reflection 2.4.4 Taking actions 2.4.5 Technology integration

The last part of the data focused mainly on SL as a platform for multimedia curriculum in terms of its technological materials, values, and defects. By using the IPA method as well, the last part of data and interpretation were extracted from a minor portion of *currere* writings and the interview transcripts of five voluntary participants (see Table 4). In the partial *currere* writings, the participants were reflecting about SL values, functionality, materials, experience, and defects. The interview portion mainly explored the participants' narratives, feelings, concept of educational application, and emotional involvement toward this technology. An attempt was made to identify the purposes of members checking for triangulation in the qualitative-based research.

Table 4. Coding scheme for technology-related data

3.1 SL Materials	3.2 SL Values	3.3 SL Defects
3.1.1 Communication 3.1.2 Subject integration 3.1.3 Multimedia integration 3.1.4 Large amount of information	3.2.1 Virtual reality 3.2.2 Emotion connection 3.2.3 Future curriculum avenue 3.2.4 Benefit for distance and diverse students	3.3.1 Time issues 3.3.2 Age-appropriate content 3.3.3 Communication issues 3.3.4 Technical issues

### The regressive stage

Pinar (2004) explicated in the regressive moment, one re-enters "past lived or existential experience as data source" (p. 36) which needs memory to enlarge and transform. During two virtual exhibition visits, the participants had time to read the information, view the images, watch the videos, and explore the related websites about the highlighted war events and ecological incidents. The data from *currere* writings and reflective notecards indicated that the exhibitions were powerful enough to bring to remembrance memories about wars or ecological events connected with their families, friends, childhood, news or lessons they learned at school, or other experiences. In the *currere* writings, the participants were asked to recall deeply their emotions, attitudes, reactions, or thoughts at the time when the specific event happened.

Through the process of entering the past, some reported being stimulated by the exhibit items, thus recalling their existential experience (regressive emotions and attitudes), enlarging their memories by thinking regressively, and generating the concurrent reflection. These steps are the *currere* regressive stages. The participants found it really useful to connect to the past experience which enabled them to process the following *currere* stages. SheriIM (avatar name), a white high school teacher, shared a paragraph from her *currere* writing.

What the lack of personal connection causes is the problem of complacency. We do not see natural disasters as a result of global warming that each person is directly contributing to. We do not see the wars as a result of foreign dependency on things like oil because we consume more than we produce. Everything is tied together in our lives but there are very few people connected enough to the problem to see the writing on the wall (SheriIM, May 10, 2011, pp. 3-4).

While narrating their autobiographies in the *currere* writings and virtual reflective notecards, some reported their emotions and attitudes were involved in deeper retrospection. Their emotions toward wars or ecological issues were released because of personal connection, morality, humanity, consciousness, and sympathy. Through the emotional expression in the regressive stage, memories were enhanced, and somehow, gained consolation.

### The progressive stage

In the progressive stage of VW activities, the participants were asked to teleport to re-created islands (see Figure 3) related to wars and ecological topics, such as the Holocaust Museum, the Vietnam Memorial Wall, Palestine-Israel War, NOAA, Etopia Eco Village and so forth. Fifteen participants specifically described being moved by the SL US Holocaust Museum for Kristallnacht exhibits which was addressed earlier. NOAA (National Oceanic & Atmospheric Administration) was another re-created island designed by the NOAA research laboratory, where participants could experience a powerful tsunami destroying a house and the melting of a glacier flowing into the ocean. The tsunami, of course, would not hurt avatars and the house would be reassembled soon after being wrecked. These virtual re-created islands simulated real-world sites in the present and actualized the imagination for a future time. Many participants reflected that they enjoyed visiting these sites because virtual reality made timelines and spaces a non-issue. Meanwhile, the participants could experience those views, buildings, materials, exhibits, and attractions with no constraint of cost or physical traveling. From aforementioned points of view, the technology of VW played a very important role, beneficial for education, regressive memories, progressive imagination, retrospection, and introspection.



Figure 3. Teleport Hubs



After visiting these re-created islands, Russcular (avatar's name), a doctoral student, shared his progressive plans: In the future, I will spend most of my time educating the public on what has been accomplished and formulating awareness tactics that will allow the public to understand how to continue to prevent these devastating situations from reoccurring (Russcular, July 2, 2011, p. 3).

### The analytic stage

The VW activities at the analytic stage were group discussion (see Figure 4) and reflective notecard writings (see Figure 5) after the participants reviewed what they had visited (background information, images, videos, webpages, and re-created environments). During the analytic stage for reviewing the past, present, and future, the participants confessed that the virtual exhibitions and the *currere* writings inspired their self-awareness of current world issues by understanding and analyzing the timeline. Some participants agreed that continuing education and lifelong learning can achieve self-awareness, and further, sublimate self-awareness toward social responsibilities both for teachers and students. Some showed their appreciation for the SL exhibition presented in the *currere* process because they had never come to that self-awareness before. As the participants travelled their timeline from the past to the future, some declared their self-awareness increased thus inspiring them to want to deliver meaningful curriculum.

Jennamb89 (avatar's name), an African American high school teacher, stated that:

My past, as I said before, was sheltered. Everything was kept from me unless it was Disney, Snoopy, or something “feel-good” related. Anything I remember was from the news and I remember the emotion I felt at the time but then it became a very distant memory. Presently, I would say I used to be complacent but now I feel informed and aware. Those things give you a sense of empowerment and that empowerment gives way to a desire for change. The biggest changes will be as I plan for my future. I feel a sense of duty and urgency to talk about issues with my students. I feel a sense of responsibility to share information that is pertinent to their future and understanding of their environment and surroundings (Jennamb89, May 10, 2011, p. 14).



Figure 4. Group discussion



Figure 5. Reflective notecard writing

### The synthetical stage

With regard to personal, global and educational points of view, self-understanding and self-reflection in the educational means of the complicated conversation encouraged the participants to synthesize the past, present, and future. Some indicated that they applied critical thinking and intellectual conversation, connecting them to their own timelines. These activities motivated them to take action to fulfill their responsibility of empowering their students to influence the world. This is the process of the synthetical stage.

Allykat18 (avatar's name), a graduate student, commented:

I learned that the past tends to repeat itself, and things that happen in the past can really shape the present and future. I always knew the past could influence the future, but I guess I never realized to what extent it could. I have learned it is important to look at topics in the context of past, present, and future—instead of just the present. You need to understand how the present came to be to truly understand it and to be prepared for the future. It is important to understand these things when going to teach students in the classroom (Allykat18, July 2, 2011, p. 15).

These participants' determination to take action was what Pinar (2004) called "self-mobilization" which occurs after self-examination, self-understanding, and self-reflexivity and it encourages them to be responsible educators for social reconstruction. Self-mobilization is aggregating self-knowledge, self-understanding, and self-reflection for reconstructing the community and the society. The ultimate goal for *currere* is to seek the "synthetical moment of mobilization when, as individual and as teachers, we enter the arena to educate the American public" (Pinar, 2004, p. xiii). This supports Pinar's point that towards the end of the synthetical stage, self-mobilization is derived and promotes social reconstruction.

## Perception of SL curriculum

SL, the main medium of virtual experience for *currere* in this study, played an important role for visual and auditory richness, multimedia integration, distance learning, considerable information, synchronous communication (see Figure 6), and virtual reality. Drawn from the data from *currere* writings and interviews, the participants shared a number of constructive perspectives about virtual SL curriculum. Regarding students' perspectives of learning about the *currere* process and curriculum development, some indicated that SL assisted students in extrapolating ideas for subject integration and interdisciplinary curriculum due to the SL demonstration of innovative installations. In the perspective of technological utilization, SL changed some of their perceptions about how integrating technology into the classroom made teaching and learning more accommodating for distant students, and further, motivated students to understand the target content more concretely and more effectively. In the perspective of autobiographic emotional involvement, SL also delivered the powerful images and videos to the participants, which directed them to understand why they possessed certain kinds of emotions towards specific events.

Issues with SL, SL Defects (3.3), is one of the important points derived from the underlying data. A few interviewees brought up the issue of insufficient time (3.3.1) for student group discussions as the allotted time was only 20 minutes for each exhibition. In addition, some participants raised the question of age appropriate content (3.3.2) in SL as they considered using this technology in K-12 teaching. SL is reserved for people who are 18 and over, which indeed is restricted to primary and secondary education. Moreover, SL has an echo problem (3.3.3) unless head sets are used by everyone. Furthermore, technical issues (3.3.4), such as internet bandwidth, permission issues, not loading video correctly, sound setup, insufficient hardware, and system overload issues, must be addressed to minimize student frustration.



Figure 6. Synchronous discussion



Russcular commented that:

Think about how many technologies you have to use in SL, such as designing your character, communicating through your microphone, adjusting your volume (to make sure you have no echo), video setting, and a lot of basic things. It is whole lot of stuff. And maybe create your own stuff, like user supplies, textures, notecards, scripts, and a website that requires a lot of innovation and technological experience. So I think SL definitely taps to creativity (Russcular, August 31, 2011).

### **Educational development of the participants**

Participants' memories, thinking, emotions, attitudes, imaginations were aroused frequently during the virtual exhibitions and *currere* writings, invoking their self-understanding, self-awareness, moral consciousness, synthetic thinking, and self-reflection. These personal syntheses followed their educational extensions—educational understanding, educational reflection, educational mobilization, motivation of educating, and social reconstruction, which involved educational application of synthetical experiences. The study indicates the process was conducive for enhancing some participants' perspectives on self-examination, educational insight, curriculum development, and technology integration. Drawn from a variety of research data, some participants indicated that they changed several of their viewpoints on delivering information, connecting to personal experience, inspiring students' interests of learning, educational meaning, pedagogical strategies, technology utilization, and teaching for benefiting the world - after SL participation and the *currere* process.

### **Conclusions**

This study significantly attempted to (1) incorporate *currere*—"the interdisciplinary study of educational experience" (Pinar, 2004, p. 2)—with the multimedia virtual world of Second Life, (2) demonstrate theoretical and practical perspectives of curriculum development to education major graduate students, and (3) identify the effects that the virtual *currere* process has on the participants' self-understanding, and take further actions for educational reconstruction. The effect of the integration initially achieved these goals. Some of participants' self-examination (e.g., autobiography, self-experience, self-understanding, self-awareness, self-reflection, and self-mobilization) and educational development (e.g., educational understanding, educational reflection, educational growth, and social reconstruction) were well processed and evolved. The suggestions for future research are twofold: (1) the development of the practical methodology to evaluate the *currere* process, realistic educational development, improvement of students' growth in consciousness, and (2) the development of interdisciplinary curricula to assess how effective the virtual *currere* approach can be.

### **References**

- Annetta, L., & Park, J. (2006). Video games in science: A model for students and teachers creating 3D role playing games. In C. Crawford, R. Carlsen, K. McFerrin, J. Price, R. Weber, & D. A. Willis (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2006* (pp. 3950-3956). Chesapeake, VA: AACE.
- Aydogan, H., Aras, F., & Karakas, E. (2010, July). *An assessment on distance education in a 3D virtual environment: How to produce electricity in a hydroelectric power plant*. Paper presented at the Education Technology and Computer (ICETC), 2010 2nd International Conference, Usak, Turkey.
- Bainbridge, W. S. (2007). The scientific research potential of virtual worlds. *Science*, 317(5837), 472-476.
- Ball, S., & Pearce, R. (2009). Inclusion benefits and barriers of "one-removed" participation. In C. Wankel & J. Kingsley (Eds.), *Higher education in virtual worlds: Teaching and learning in Second Life* (pp. 47-64). Bingley, UK: Emerald Group Publishing Limited.
- Barab, S., Thomas, M., Dodge, T., Carteaux, R., & Tuzun, H. (2005). Making learning fun: Quest Atlantis, a game without guns. *Educational Technology Research and Development*, 53(1), 86-107.

- Boulos, M. N. K., Hetherington, L., & Wheeler, S. (2007). Second Life: An overview of the potential of 3-D virtual worlds in medical and health education. *Health Information & Libraries Journal*, 24(4), 233-245.
- Brown, D. S. (2007). *Restorying ourselves: Using currere to examine teachers' careers*. Unpublished doctoral thesis, Indiana University, Indiana, USA. Retrieved from <http://digital.library.okstate.edu/etd/umi-okstate-2383.pdf>
- Clark, S., & Maher, M. L. (2001). The role of place in designing a learner centred virtual learning environment. *Proceedings of the Ninth International Conference* (pp. 187-200). doi: 10.1007/978-94-010-0868-6\_15
- Cheal, C. (2007). Second Life: Hype or hyperlearning? *On the Horizon*, 15(4), 204-210.
- De Freitas, S., & Martin, O. (2006). How can exploratory learning with games and simulation within the curriculum be most effectively evaluated? *Computers and Education*, 46(3), 249-264.
- De Freitas, S., Rebolledo-Mendez, G., & Liarokapis, F. (2010). Learning as immersive experience: Using the four-dimensional framework for designing and evaluating immersive learning experience in a virtual world. *British Journal of Educational Technology*, 41(1), 69-85.
- De Lucia, A. D., Francese, R., Passero, I., & Tortora, G. (2009). Development and evaluation of a virtual campus on Second Life: The case of SecondDMI. *Computer & Education*, 52(1), 220-233.
- Dickey, M. D. (2005). Three-dimensional virtual worlds and distance learning: two case studies of Active Worlds as a medium for distance education. *British Journal of Educational Technology*, 36(3), 439-451.
- Dillenbourg, P., Schneider, D. K., & Synteta, P. (2002). Virtual learning environments. In A. Dimitracopoulou (Ed.), *Proceedings of the 3<sup>rd</sup> Hellenic Conference "Information & Communication Technologies in Education"* (pp. 3-18). Rhodes, Greece: Kastaniotis Editions.
- Ditcharoen, N., Naruedomkul, K., & Cercone, N. (2010). SignMT: An alternative language learning tool. *Computers & Education*, 55(1), 118-130.
- Eatough, V., & Smith, J. (2006). 'I was like a wild wild person': Understanding feelings of anger using interpretative phenomenological analysis. *British Journal of Psychology*, 97(4), 483-498.
- Erra, U., & Scanniello, G. (2009). Synchronous communication media in the software requirements negotiation process. *Cooperative Design, Visualization and Engineering*, 5738, 245-252.
- Gollub, R. (2007). Second life and education. *Crossroads*, 14(1), 1-8.
- Graves, L. (2008). A Second Life for higher ed. *US News & World Report*, 144(2), 49-50.
- Greenhill, K. (2008). Do we remove all the walls? Second Life librarianship. *Australian Library Journal*, 57(4), 377-393.
- Grumet, M. R. (1976). Psychoanalytic foundations. In W. F. Pinar & M. R. Grumet (Eds.), *Toward a poor curriculum* (pp. 111-146). Dobuque, IA: Kendall/Hunt.
- Inman, C., Wright, V. H., & Hartman, J. A. (2010). Use of Second Life in K-12 and higher education: A review of research. *Journal of Interactive Online Learning*, 9(1), 44-63.
- Irwin, R. L. (2006). Walking to create an aesthetic and spiritual currere. *Visual Arts Research*, 32(1), 75-82.
- Jarmon, L., Traphagan, T., Mayrath, M., & Trivedi, A. (2009). Virtual world teaching, experiential learning, and assessment: An interdisciplinary communication course in Second Life. *Computers & Education*, 53(1), 169-182.
- Kanade, T., & Rander, P. (1997). Virtualized reality: Constructing virtual worlds from real scenes. *Multimedia, IEEE*, 4(1), 34-47.
- Kluge, S., & Riley, L. (2008). Teaching in virtual worlds: Opportunities and challenges. *Issues in Informing Science & Information Technology*, 5, 127-135.
- McComas, J., Pivik, J., & Laflamme, M. (1998). Current users of virtual reality for children with disabilities. In G. Riva (Ed.), *Virtual environments in clinical psychology and neuroscience: Methods and techniques in advanced patient therapist interaction* (pp. 161-169). Amsterdam, Netherlands: IOS Press.
- McKay, S., Van Schie, J. & Headley, S. (2008). Embarking on an educational journey in Second Life. In K. McFerrin, R. Weber, R. Carlsen, & D. A. Willis (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2008* (pp. 1762-1766). Chesapeake, VA: AACE.

- Minocha, S., & Reeves, A. J. (2010). Design of learning spaces in 3D virtual worlds: An empirical investigation of Second Life. *Learning, Media and Technology*, 35(2), 111-137.
- Pinar, W. F. (2004). *What is curriculum theory?* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Roussou, M. (2009). A VR playground for learning abstract mathematics concepts. *IEEE Computer Graphics and Application*, 29(1), 82-85.
- Sameshima, P., & Irwin, R. (2008). Rendering dimensions of a liminal currere. *Transnational Curriculum Inquiry*, 5(2), 1-14.
- Smith, J. A., & Dunworth, F. (2003). Qualitative methods in the study of development. In K. Connolly & J. Valsiner (Eds.), *The handbook of development psychology* (pp. 603-621). Thousand Oaks, CA: SAGE Publications Inc.
- Smith, J. A., Jarman, M., & Osborn, M. (1999). Doing interpretative phenomenological analysis. In M. Murray, & K. Chamberlain (Eds.), *Qualitative health psychology* (pp. 218-240). Thousand Oaks, California: SAGE Publication Inc.
- Thompson, S. (2008). Teaching in a virtual setting. In L. Bell, & R. B. Trueman (Eds.), *Virtual worlds, real libraries: Librarians and education in Second Life and other multi-user virtual environments* (pp. 165-172). Medford, NJ: Information Today, Inc.
- Vrellis, I., Papachristos, N. M., Bellou, J., Avouris, N., & Mikropoulos, T. A. (2010, July). *Designing a collaborative learning activity in second life-an exploratory study in physics*. Paper presented at the 2010 10th IEEE International Conference on Advanced Learning Technologies, Sousse, Tunisia.
- Wang, Q., & Woo, H. L. (2007). Comparing asynchronous online discussions and face-to-face discussion in a classroom setting. *British Journal of Educational Technology*, 38(2), 272-286.

## Appendix A

### Stage One – Regression:

Do you remember any wars or ecological details that occurred in your childhood or youth?

☐ Yes ☐ No

What wars or ecological details do you remember? Please be specific about the details, including places, time periods, governmental reactions, industrial reactions, educational responses, citizen responses, impacts, etc.

Describe one example of wars or ecologies from your school that stands out.

Why are your memories of the occurrence still vivid?

Were these two Second Life exhibitions strong enough to make you recall people or events related to the wars and ecological occurrences you are familiar with?

☐ Strongly Disagree ☐ Disagree ☐ Agree ☐ Strongly Agree

What parts/aspects of the exhibitions stimulated the recall?

After seeing the exhibitions, what are your thoughts, reflection, or understanding about wars and ecological issues?

### Stage Two – Progression:

Imagine that the world will be devastating in the 22<sup>nd</sup> century as a result of climate change, lack of food and water, and wars between countries. What are you going to do as an educator to prevent these tragedies from happening?

In what way and with what tools will you deliver the curricula to your students regarding wars and ecological issues?

Imagine you are about to receive an award in recognition for your contribution for delivering a tremendous curricula regarding wars and ecology. How will you feel about it?

Imagine you successfully prevented the devastating situations as predicted by scientists, what will be your next set of goals for next five to ten years?

### Stage Three – Analysis:

As you ruminate over the past, present, and future, what ideas show up over and over again?

As you recall the past, what parts do you wish to be done differently?

What parts make you sad?

What parts make you angry?

What parts leave you with mixed feelings?

3. What decisions did you make today as a result of influences from your previous responses? In what way(s) were decisions you made today influenced by your hopes for the future?

4. As a pre-service or in-service teacher, what are your thoughts at this stage of analysis stage?

What are the big ideas or themes you see in your professional life?

### Stage Four – Synthesis:

Was what you believed about yourself confirmed through this process? If yes, what are they? If no, what are they?

What self-understanding (self-knowledge and self-belief) did the virtual exhibitions make you come up with?

What self-reflection did virtual exhibitions make you come up with?

What have you learned from peers and classroom discussions regarding virtual exhibitions?

What are your thoughts on using virtual world to present an interdisciplinary curriculum?

What have you, as a pre-service or in-service educator, learned from the synthetic process of the past, present, and future?

If you are an interdisciplinary instructor and you had to design a curriculum about globally persistent issues, what themes will you come up with? What methods will you use to deliver the curriculum? Why did you choose this topic to present?

## **Appendix B**

### **Interview Protocol**

Do virtual exhibitions impact your perceptions about teaching and learning a curriculum?

What are your thoughts about using Second Life to deliver a target curriculum? Could you please respond this question in the perspectives of technology, creativity, and effectiveness?

What are your thoughts about the use of communications (text chat, voice chat, public chat, and private chat) to engage distance complicated conversation?

What do you think you learned from the Second Life component of this course that you would not have learned otherwise?

For what other types of learning activities do you think Second Life would be potentially effective?

Did you observe any change(s) in your attitude towards teaching and learning after participating in Second Life activities?

# Remote Laboratories: Design of Experiments and Their Web Implementation

Ananda Maiti\* and Balakrushna Tripathy

<sup>1</sup>School of Computing Science and Engineering, VIT University, Vellore 632014, India // anandamaiti@live.com // tripathybk@vit.ac.in

\*Corresponding author

(Submitted March 17, 2012; Revised June 03, 2012; Accepted July 16, 2012)

## ABSTRACT

This paper discusses the technical considerations and pedagogical issues that influence the design and classification of hardware-based experiments and their implementation in an internet-based remote laboratory environment for technical education. The necessary architecture to maintain academic integrity and imparting quality laboratory education via remote laboratories involving the instructor, faculty, and the students has been discussed. Online laboratory teaching and learning may be augmented by adding voice/multimedia protocols for teacher-student interaction. Pedagogical issues conducive for online laboratory instructions incorporating effective remote learning facilities that mimic the face-to-face interaction in conventional laboratory have been discussed. An online remote laboratory management system ensuring quality of service, security, automated laboratory report evaluation, and monitoring of student progress has been proposed. The main contribution of this paper is to develop a reference model which may help the educators and hardware-based laboratory module developers for electrical engineering disciplines.

## Keywords

Remote laboratory design, Classification of experiments, Scheduling, Web services, Pedagogy

## Introduction

The importance of hands-on experimentation in the undergraduate electrical engineering curricula is well established. A remote laboratory provides a web-based platform to the students to access, control and perform experiment from remote locations. Increasingly electrical engineering departments are beginning to incorporate remotely operated hardware-based laboratories in laboratory curricula. Several institutions have also developed remote laboratories as a supplement to their existing laboratory infrastructures. Provided the experiments are set up to run without the need for human intervention, remote laboratories can be made available via the Internet on a 24x7 basis, facilitating access at all times that accommodate individual student's needs. It is a group effort, where content authors (educators), instructional designers, multimedia technicians, database administrators, and people from other areas of expertise come together to serve the community of learners. To date, very few studies have been performed to determine the special requirements needed for the design of laboratory modules consisting of several experiments and their web implementation and the effectiveness of the designed experiments towards the fulfillment of the educational goals of traditional laboratories.

In this paper, first a brief overview of state-of-the-art technologies in the development of remote hardware-based laboratories is given. The development of a remote laboratory module faces several challenges; in design and implementation of the experiments for web launching, the online laboratory learning environment design that ensures strong, effective, accessible, and secure student interaction that best replaces the face-to-face (F2F) interaction that takes place in conventional laboratories. The paper is organized as follows. Section II presents a brief review and the desired pedagogical features of remote laboratory education. Section III discusses how the traditional laboratory experiments can be modified and tailored for online delivery. Section IV describes various experiment classification schemes. Section V discusses the instructor's role in laboratory management system. Section VI describes the laboratory architecture. Section VII discusses the implementation issues of a hardware-based remote laboratory module and authentication and access control issues. A case study is discussed in Section VIII. Section IX presents the results of evaluation of the remote laboratory followed by conclusions and future directions.

## Pedagogical features

Online laboratories are classified into remote experiments (based on actual physical setup) and virtual experiments (based on simulation). In majority of cases, the remotely accessible online laboratories reported in the literature are dispersed around the world and represent stand-alone systems. Hardware-based remote laboratories are being successfully used in electrical engineering education. Several remote laboratories in which the learners interact and operate the hardware from a remote location are now operational at various universities (Harward et al., 2008; Fjeldly, & Shur, 2003; Gomes, 2007; Auer, 2003; Trevelyan, 2004; Maiti, 2010a). Internet laboratory courses covering many other disciplines of engineering and science are being offered to students worldwide. A few examples include MIT iLab, European Schoolnet, and PROLEARN, besides many other individual and/or collaborative distance laboratory efforts all over the world. Recently, the world consortium Library of Laboratory (LILA) has been formed to furthering the use of the remote laboratory in engineering education.

Several interesting reviews on remote electronic laboratories have appeared (Sivakumar, 2005; Campbell, 2002). Guimaraes et al. have discussed the design and implementation of remote laboratories (Guimaraes, 2011). State-of-the-art technologies used in the development of remote laboratories have been presented by Gomes and Bogosyan (Gomes, 2009). A special mention is due to the Special Issue recently published in the IEEE TRANSACTIONS ON EDUCATION devoted to virtual laboratories (Reilly, 2008 and references there in). While these reports provide comprehensive discussions of a wide variety of technologies for web-based remote laboratories, most journal or conference articles focus on one particular laboratory experiment (from different disciplines) and emphasize mostly on the educational and/or technological aspects. However, important issues such as, the development of a remote laboratory module consisting of several experiments, management of experiments, quality of service, security, safety and the operation of the equipment have not been addressed adequately. The design of remote hardware-based laboratory experiments and their web-implementation present different levels of complexity depending on the nature of the experiment and equipment necessary for running the experiment. For example, the development of a fully remote controlled laboratory module (consisting of 10 experiments, say) experience goes well beyond the design and implementation of a single experiment. The development of a remote laboratory module goes well beyond the design of the experimental apparatus and support software. Classification and selection of experiments (for example, a short or a long duration experiment) itself is a big issue in developing the experiment module. The design of experiments must be pedagogically sound to ensure that students achieve the same level of performance (and experience) with respect to the learning outcomes as would be expected from a conventional laboratory.

Remote laboratory module development needs to address the specific issues pertaining to pedagogy, facilitation, scalability, usability, and security within a framework. Conventional laboratory development procedures cannot be directly employed and need modifications for remote laboratory practice for the students. During the development of instructional materials and delivery methods, the developers need to address these issues, like; technical infrastructure management for delivering learning material; curriculum quality; system scalability to handle increases in student enrollment; and continuous student assessment for grading purposes. Towards the development of student-centric laboratory teaching process, the design of the pedagogical strategy should include; user-friendly system interface; provide real-time response from equipment; provide a collaborative learning environment for group interactions; to engage students actively in achieving learning outcomes; and track student performance to meet learning outcomes. From the technology-centric viewpoint, the instructional delivery framework should address several issues, like; use standard networking practices and protocols to connect the user to the server/equipment; ensure security; and track student access to laboratory resources. Also, for instructors, remote laboratory educational models require far more work than traditional laboratory educational models, currently being followed. Generally, traditional face to face (F2F) laboratory teaching professors are not in a position to deliver successful remote laboratory courses due to the lack of their pedagogical training. Instructors generally find it difficult to change their delivery methods. For successful deliver of remote laboratory courses, teacher training is necessary.

Remote laboratories also need to be managed with a laboratory management system (LaMS). The primary objective of the LaMS is to manage the learner database, keeping track of their progress and performance across various stages and types of training activities. The LaMS should also be able to manage student registration, scheduling of experiments, and allocate resources such as the instructor availability, serve instructional materials, delivery of remote learning materials, and authoring system. The LaMS needs to provide a scheduling system that ensures that each user or a group can use the hardware without disturbance during experiment. The LaMS should be able to save

the experimental data, retrieve and evaluate and compare data on learner scores, navigation habits and use them to provide the content managers (educators) the crucial information on the effectiveness of the content, if combined with specific instructional strategies, delivery technologies, and learner preferences. It should also provide an infrastructure that can be used to create, modify, and manage contents for a wide range of learning environments to satisfy the needs of rapidly changing laboratory education requirements. Also, if needed, LaMS may be used to extend remote laboratory facilities across multiple institutions through shared resources. By sharing remote laboratories resources, high cost involved in developing hardware-based laboratories can be significantly reduced.

## **Design of experiment for remote laboratory**

A traditional laboratory experiment setup requires a substantial re-working before it can be launched on the web. The steps to convert a traditional laboratory experiment setup into a remote online version include the following design and pedagogical issues:

- The first step is to identify the type of experiment. Experiments may vary widely from each other depending upon their domain or field. Various schemes for classification of experiments will be discussed in section IV.
- The second step is to identify a proper scheduling algorithm based on the type of experiment. Since users do not know about other users actions, the laboratory server should have a mechanism to handle users' requests free and fairly.
- The third step is to decide on the switching/change over that may be required in an experiment. In a remote laboratory, the students need to change the experimental setup and connections. For example, one can use robot for automatic change over and the robots need to be controlled through web services and their control must be added to the user interface in step 4.
- The fourth step is to design the equipment interface. There are several ways to design the interface such as, browser based application or a stand-alone application. For the laboratory interface to be more robust and portable, it is always desirable to create the interface in a browser based script or language. Several drag and drop IDE such as Adobe Flex are available to create attractive and lightweight application that can be easily run through the browser on any platform. The user interface calls the web service that runs the experiment on the server side.
- The fifth step is to select the data exchange protocol. Different experiments may generate of different types of outputs. The interface needs to call a specific web service for a specific experiment and handle the output uniquely.
- The sixth step is to decide on the security measures. The most important is to ensure that no arbitrary request is processed which could damage the instruments and the access to the instrument must be highly secured.
- If desired, the server may be provided with a monitoring system using multiple cameras to show users the functioning of the instruments during the experiment.
- Although a remote laboratory may be run in a 24x7 mode, the equipment may be put 'on' only during experiment. Power control units may be used to turn 'off' and 'on' the instruments through a thread or program which will run in the server. The program will turn 'on' the instrument just before it is to be used and will 'shut down' when the experiment is over.
- The remote laboratory should be made, as far as possible, to mimic the conventional laboratory. This may be done by adding multimedia elements such as, videos, images, and manuals for each experiment.
- An interactive animation may be used to explain the experimental setup to users through a series of scenes showing how the experiment was setup and connections are made. Towards interpretation of experimental data, additional resources may also be provided.
- Nowadays, blogging and social media has become a new tool for learning among peers. This may also be applied in a remote laboratory by creating forums and video conferencing. The forums can be used to post questions which are answered by the teachers. If required, students and the teacher may have a video conferencing session to discuss on various experiments.

## **Schemes for classification of experiments**

Remote hardware-based online Microelectronics and VLSI Engineering laboratory (NETLab) has been designed and developed by the Indian Institute of Technology (IIT) Kharagpur and VIT University, Vellore (Maiti, 2010a). The



laboratory is operational for use in the undergraduate and postgraduate laboratory teaching for more than two years now and has served more than 300 students. Currently available six experiments are listed in Table 1 which includes the following attributes: (a) experiment name, (b) measuring equipment, (c) input parameter for measurement, (d) output data from the experiment, (e) typical measurement setup time, and (f) actual measurement time. Primary classification of experiments may be made from these attributes, for example, experiment on MOS capacitor is a ‘short’ experiment while Atomic Force Microscope is a ‘long’ experiment and are described in detail below.

SL No.	Experiment Name	Measuring Instrument	Input Variable	Output Variable	Setup time	Actual Measurement Time
1	MOS Capacitor	Agilent E4980A	V, F	V,I as data set/array	5 min	1-2 minutes
2	BJT Characterization	Agilent 4156C	V	V,I as data set/array	10 min	15-20 seconds
3	BJT Gummel plot	Agilent 4156C	V	V,I as data set/array	10 min	20-25 seconds
4	MOSFET ID-VD Characterization	Agilent 4145B	V	V,I as data set/array	7 min	15-20 seconds
5	MOSFET ID-VG Characterization	Agilent 4145B	V	V,I as data set/array	7 min	20-25 seconds
6	Atomic Force Microscope	NanoSurf	Set of numeric values for area and scan lines	Image	20 min	15-20 minutes

F: Frequency V: Voltage I: Current

Table 1. Partial list of experiments available in the Microelectronics and VLSI Engineering laboratory.

In the following, we classify the experiments based on the attributes shown in Table 1 and describe briefly their implementation in the LaMS to be described in Section VII.

- **Classification based on input parameters:** In this regard experiments can be classified into two categories: (a) all experiments require inputs, for example, MOS capacitor requires two input parameters viz., voltage range (say -3V to +3 V) and the measurement frequency (say, 1 MHz). Typically, this may be considered as ‘small input’ (in terms of bytes consumed) as the size limit could be within 150-160 characters. In fact, such experiments can be run even by sending Short Message Service (SMS) from a mobile and (b) ‘Large Input’: This type of experiments require bulk data. The number of parameters is usually large and there is no restriction on how many characters the input may be composed of. For example, Atomic Force Microscope experiment requires a 2-D dataset as input for the area on which scanning is to be carried out.
- **Classification based on output parameters:** Based on the nature of output data generated by the experiments, they can be classified into two categories: (a) ‘large output’ where the experiments generate huge amount of data which is generally raw data and does not convey any meaning unless drawn as a graph or plotted as a chart. For example, BJT characterization experiment generates x-y (V-I) dataset which needs to be plotted for interpretation, (b) Continuous: some experiments may keep generating data and the server stores them. The user needs to specify the time slot (say up to 10 sec) for which they wish to take data. Once such a request is received, the server creates a thread that fetches the necessary data from the database and sends it back to the server.
- **Classification based on processing time:** For each experiment, there will be; (a) time required for setting up the measurement conditions, and (b) experiment real run time. ‘Short’ experiments are classified based on the criteria that both the setup and real run time are small. For example, MOSFET characterization experiment setup takes about 5 minutes and run time is only 15-20 seconds. ‘Medium’ experiments take longer time compared to short

experiments but much less time compared to ‘long’ experiments such as Atomic Force Microscope experiment which takes almost 20 minutes for setting up and scan time could be as high as 30 minutes. The above classification helps in many ways to set the ‘schedule’ for each experiment which in turn is used for time-slot fixing for an experiment, queuing etc. Both the SHORT-MEDIUM-Scheduler and LONG-SCHEDULER have been developed (Maiti, 2011a; Maiti, 2011b). SHORT-MEDIUM-Scheduler may be used to run experiments in a multi-user mode i.e., more than one user can operate the instrument in an allotted time slot. An experiment request arrives at the server through a HTTP request or web service request. After the request is received all setup parameters are verified and then the experiment is triggered.

- **Classification based on interactivity:** There are two types of experiments; interactive and non-interactive. Some experiments may require constant interaction with the interface; some may not require any, while some may require constant feedback throughout the duration of the experiment. Experiments that take some inputs from the user while running are definitely single user, i.e., one user must be given a dedicated time slot for the experiment. This leads to the implementation of complex scheduling schemes. Experiments which run without an intermediate user interaction can be implemented by ‘long’ or ‘short’ scheduler as discussed earlier.

### Instructor’s role in laboratory management system

The instructor is the administrator of the remote laboratory module who creates and maintains the learning contents. All instructor functions are performed online through the LaMS which menu driven. The followings are few of the instructor’s functions performed online:

- **Create/alter experiment contents:** Creating an experiment involves adding the basic details of an experiment such as aim, procedure, and description, expected duration etc. The instructor also uploads the learning materials such as the laboratory manuals, videos animations and additional reading materials through the online system. Finally, the experiment interface is uploaded in the system.
- **Create viva questions/evaluation of laboratory reports/student records:** For the evaluation of students understanding viva is conducted online and laboratory reports are also submitted online by students. The instructor uploads viva questions and expected (default) results for each experiment in the database. The student’s laboratory report evaluation has also been automated and submitted results are compared to the experiment’s default data and grading is done. Grades are then uploaded automatically for the students and also for the instructor.
- **Question box/video conference:** The questions posted by the students for an experiment are answered by the teachers and if required, video conferencing may also be arranged for the students.
- **Cameras:** Each experiment setup may be associated with one or more cameras that stream the screen of the instruments during setup and measurement. The video streaming is done through webcam connected to the server. The Flash Media Server is used for the video streaming. A unique key is associated with a camera that is embedded or passed to the user interface depending upon the experiment being performed.
- **Keeping track of defaulters:** The remote laboratories are designed for increasing the efficiency of laboratory education. If a student books a time slot and fails to conduct the experiment during that time, the system records it so that the instructor can advise the student a new time slot.

### Laboratory architecture

The general architecture of NETLab is shown in Figure 1 and is described as follows:

- **Local or instrument server:** The server controls the measuring instruments. The server hosts the actual web service and other services as well. Each instrument is connected to a power control unit (PCU) which turns the instrument ‘on’ and ‘off’ and also resets them.
- **NETLab server:** The main NETLab Server takes all the user requests and processes the requests. It invokes the local server’s services for triggering the experiment. It also manages the experiments, user’s details and the instruments. The system is implemented using Java and Oracle technologies. The NETLab Server also have

provisions for user management such as evaluating the student performance and recording their progress online (Maiti, 2010b)

- **Users side:** The Client Side consists of a Web browser interface and is platform independent. The interface is used to provide the input parameters for an experiment and display output data as graphs, charts or tables.

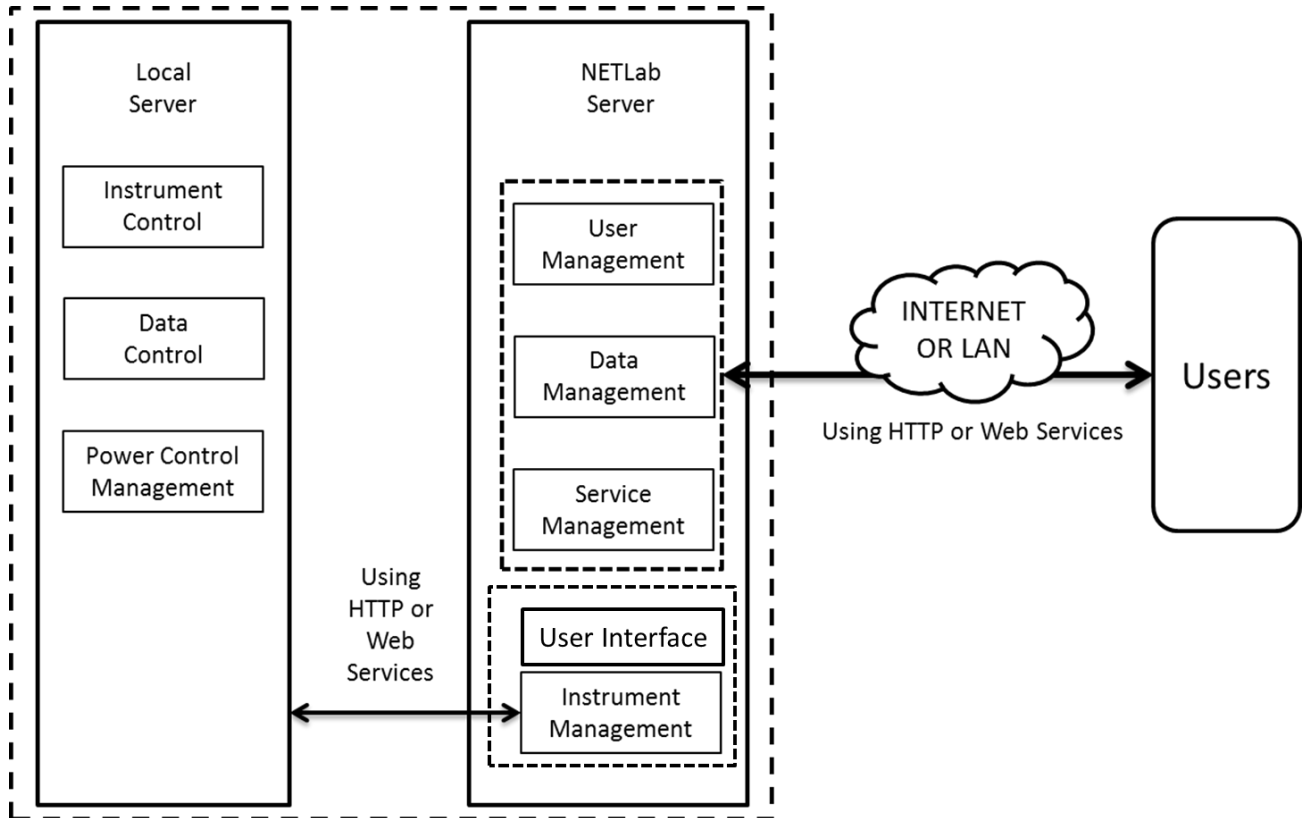


Figure 1. The NETLab Architecture.

Experiments are run after receiving a request from the client interface. A typical experiment run (see Figure 2) may be described as follows:

1. *Input:* The user submits the input as a set of parameters. The input is taken from the user interface via a web browser.
2. *Processing of inputs:* The processing of inputs parameters is done two steps:
  - *Validation:* An input may be wrong i.e., not relevant to the experiment resulting in erroneous output. Also the input parameters may be faulty which may cause damage to the instrument. In the validation phase, checking of the input for such patterns are done and also whether the experiment can be performed with the given inputs.
  - *Saving data:* This is an optional step. For an experiment involving large input datasets, the data need to be saved on the server's hard disk before starting the experiment. The saved data need to be in a particular format and attributed to the user for performing the experiment.
3. *Performing the experiment:* An experiment run involves three steps:
  - *Triggering the experiment:* Every experiment needs to be associated with an http link such as, a web service. The actual experiment may not be performed at the main server but instead at a local or instrument server associated with the real instruments. The input is passed as GET or POST parameters to the web services. The instrument server then starts the experiment.
  - *Carrying out the experiment:* Since instruments, involved with the experiment, need to be 'ON' for entire duration of the experiment, the instruments need scheduling such that no two or more instances of the experiment are initiated at the same time. For this, the instruments are associated with tokens. This step has to collect all

tokens for the experiment such that no other experiment can gather them. This involves changing the data in the database associated with the instrument.

- *Completing the experiment:* After the experiment is completed, the data is gathered from the instrument and sent to the main server. The token associated with experiment is now released i.e., their values are reset in the database such that the next experiment can commence.
4. *Error Reporting:* This is an optional step and is required, if and when an error occurs, i.e., if an instrument fails to generate the correct data. The error has to be captured and identified. The user is notified about the failure and immediate corrective action is taken so that the experiment run can be resumed.
  5. *Output:* The last step in an experiment is to send the output data to the user. Since the request is sent over http, the response is also over http. If for certain experiment, the output is large, then it may be saved along with the users' details on the server for sending it to the user later.

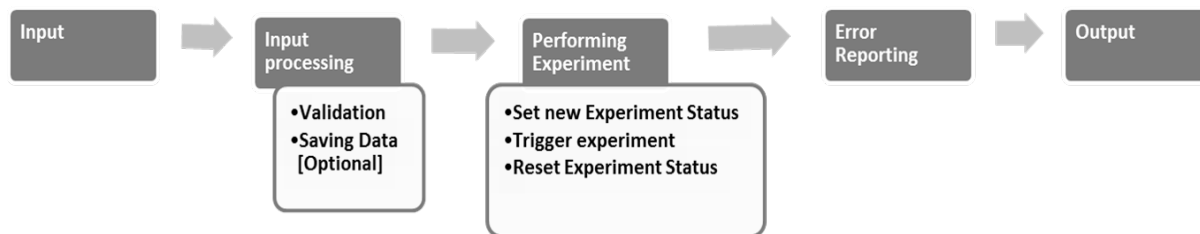


Figure 2. Experiment Process or an 'experiment run' when a user's request arrives.

## Implementation

The remote laboratory experiments require an extensive implementation planning. Although each experiment is different, but they share common input output and data exchange protocol. The three most important issues in implementing an experiment are:

### Methods of data exchange:

The data has to be transferred from the client to the server and then back to the client. The best way to deal is to use a set of web services. A web service is a block of code that is situated on the server and invoked from remote computer through the web. Web services are in two flavors: REST and SOAP. For implementing the experiment both of these are suitable. A web service represents an experiment or in some cases an instrument. The remote user client calls the web services with the given inputs as parameters. Web service technology is provided with its own security and network management protocols.

Grid computing is a collection of computer resources from multiple domains that work together to reach a common goal. In remote laboratory applications, a grid computing architecture in which the laboratory is divided in 3 levels forming 3-tier architecture may be followed. The lowest level is the local server. The second level is the main (NETLab) server. This server controls all local servers simultaneously to start, execute, collect data and stop the experiment. The grid architecture allows the laboratory to be more flexible since the laboratory in itself can be spread over a large geographical area. Secondly, it ensures reliability and robustness, since even if one node fails, the main server can run and serve the clients.

Role of cloud computing environment in the implementation of remote laboratories is being explored (Saliyah-Hassane et al., 2011). It has been proposed that based on the concept of 'Software as a Service' (SaaS) will allow the students to use software tools from the Cloud to perform the tasks in solo or in collaboration. SaaS is a scheme in the cloud computing which provides the functionalities of a software as a set of services. The user calls these services

and the service provider decides where to execute the request and once the request is processed, the output is sent back to the user. The entire processing is distributed over several locations optimizing the process and hiding it from the user. Virtual (simulation-based) laboratories may get best benefit out of cloud computing as virtualization enables one server or computer to act as many. The benefits of bringing a cloud computing approach is to use the laboratory resources, where students are able to use virtual machines so that a student's computing resources can be accessed from anywhere and can be considered as truly portable.

### Format of data exchange:

The second important issue is the format of data that is generated by an experiment. An experiment is expected to generate data in a particular style only. For example, the experiment Gummel Characteristics will generate a plot containing two curves, i.e., there will be two Y-axes and one X-axis data. Similarly, there may be experiments where the data is generated in tabular form and the number of columns or rows may not be the same. Hence, one needs a standard but flexible method of representing data, such that the data generated by an experiment is understood by all parties i.e., the software part in the entire application. Such functionalities are well presented in eXtended Markup Language (XML). XML (see Figure 3) can represent any kind of data with ease and clarity. Also, with XML there is platform independence since the XML can be parsed by almost all major languages.

```
<ARRAY>
  <DBL>
    <VALX>0.60</VALX>
    <VALY1>0.001</VALY1>
    <VALY2>0.000</VALY2>
    <VALY3>-4.86E+10</VALY3>
    <VALY4>-6.93E+09</VALY4>
    <VALY5>0.000E+00</VALY5>
  </DBL>
</DBL>
```

Figure 3. XML based structure of data exchange.

### Security of the laboratory:

Security of the laboratory involves two major areas:

- *Data Security*: As the remote laboratory is used mostly for educational purposes, it does not generate any kind of confidential data. The only security for data required is the student's information stored in the system database. This part is similar to any other information system storing user information. Apart from this, the identifier or token to represent each instrument have to be secured on the server, since these can be misused for unauthorized use of the instrument and possibly for disrupting the system. It should also be renewed or changed from time to time to prevent risk from accidental loss or theft of data.
- *Instrument Security*: The more important part of security in remote laboratories is the instrument security. The instruments in a remote laboratory are handled by students with little knowledge about the instrument. Hence the instruments are at a greater risk of getting damaged. In an on-site laboratory, instructors check the student's actions and stop possible hazards. In a remote laboratory, the management system must provide adequate measures to guarantee the instrument safety. Only way to ensure it is to prevent any malicious input being processed. This is done at the second step in experiment process to validate the inputs.

Instrument sharing may be done in three different levels:

- *Inflexible or static sharing*: Usually an instrument has many possible input areas such as 'buttons', 'knobs' etc. A combination of these inputs and their function is required for a particular experiment. This combination is unique for an experiment. At this level of security makes sure that the instruments are used only in some specific ways

i.e., only a few specific functions are implemented over the web and the instrument cannot be used in any other mode. The functions implemented need to be specific for an experiment.

- *Flexible or Partial Sharing*: In this case, input from a button or a knob is associated with a corresponding function. Various combinations of the function can be used to perform to set a certain range of values for an experiment. However, the more controls are given to users will increase the chance of more invalid inputs requiring a stricter validation scheme.
- *Complete or total sharing*: In total sharing, each of the input nodes is associated with a function that is available over the network along with the capability to turn ‘on’ and ‘off’ the instruments. The user has the total control of the equipment. Although this is necessary for learning point of view, but requires an extremely strict and strong validation scheme to validate the inputs.

It may be noted that with the increase in flexibility of controls for the users over the instruments, safety risks also increases. Hence, while implementing a remote laboratory, the administrators must decide a balance of flexibility and the instrument security.

### **Operability with mobiles devices:**

Now-a-days mobile devices such as, cell phones and smart phones have become very popular with the students community. Also the charge of using the internet on such devices along with the use of messaging services such as, Short Message Service (SMS) are efficient, easy to use and reliable. Design and implementation of an online laboratory system in flash-enabled mobile platform such as, pocket PC, smart phone, and PDA using FLASH Lite 2.0 has been reported (Maiti, & Tripathy 2011c), (Maiti, 2010c). The Microelectronics and VLSI Engineering Laboratory (<http://lod.iitkgp.ernet.in/netlab>) described in Section VIII may also be operated using smart phones which provides greater flexibility in student learning as one need not to be at home or laboratory to perform the experiments.

### **Case study**

The NETLab, described in Section IV is operational for use in the undergraduate and postgraduate laboratory teaching at the Indian Institute of Technology (IIT), Kharagpur. The NETLab laboratory management system (Maiti, 2010a) is being used to implement different types of hardware-based online laboratory modules; such as, Solid-State Devices Laboratory (<http://ssd.iitkgp.ernet.in/netlab>) and technology CAD (TCAD) laboratory (<http://tcada.iitkgp.ernet.in/netlab>). In the following, detailed operation of a remote experiment (Gummel plot) available in the hardware-based Microelectronics and VLSI Engineering laboratory is described in detail.

The students perform the experiments in the following steps:

#### **1. Preliminary quiz:**

It is necessary that the student engages the equipment only when the experiment is performed. A preliminary quiz has been introduced and conducted before the learner can get access to the actual experiment webpage and the equipment and/or the experimental setup. This is done primarily to optimize the use of the instruments by making sure that the user is familiar with the instrument and knows about the experiment. The student gets access to the experimental setup only during the time slot booked.

#### **2. Booking of time slot:**

The user must book a time slot for performing an experiment. For user’s convenience, available time slot is highlighted for 15 days. If ‘queuing’ scheme is applied, only one user can book a time slot. However, if a ‘hybrid scheduling’ scheme is operational (Maiti 2011b) more than one user may book the same time slot.

### 3. Running the experiment:

The user performs the experiment by opening the experiment user interface, setting the desired measurement parameters and then triggering the experiment. Upon completion of the experiment, the results/data are displayed on the user screen. The result may be copied from the interface and stored for analysis by the student in his/her PC.

We shall emphasize on the bipolar device characterization experiment which makes use of a single instrument (Agilent 4156C Semiconductor Parametric Analyzer) for the characterization of a bipolar junction transistors (BJT) involving five experiments sequentially. The instrument is controlled by LabVIEW in combination with data acquisition cards (Travis, & Kring, J.2006). A local server is assigned to operate on it. To increase power efficiency a Power control unit (PCU) is used that supplies power to the instrument. It is connected through the serial ports to the computer. The computer can send commands to the PCU to turn 'on' or 'off' the instrument. The Experiment Module for BJT characterization consists of the following five experiments (with code):

- BJT Gummel plot (s1002)
- BJT Output Characteristics (s1003)
- Collector Resistance (s1013)
- HFE Characteristics (s1014)
- Emitter resistance (s1015)

Both the instrument and PCU are controlled i.e., their functions are invoked using RESTful Web Services created by using LabVIEW. The LabVIEW files are designed to set input parameters on the instruments, run the experiments, and acquire data from the instrument. After this, the data is converted to XML format. When the NETLab server receives a request from the internet first calls the PCU web services followed by the instrument web services and finally return the result to the client making the request. Using this architecture several different instruments located at different sites can be controlled and operated by the NETLab laboratory management system.

One of the most important experiments in evaluating bipolar transistor parameters is measuring the collector and base current as a function of base-emitter voltage (Gummel plot). The measured data can be graphically analyzed to obtain saturation current, current gain, current gain vs. collector current characteristics, along with base resistance and recombination current characteristics. The user performs the experiment by opening the experiment user interface, set the desired measurement parameters. For example, for running this experiment, the user needs to set the channel definitions as shown in Figures 4 (a-c) where the measurement conditions (voltage and current) are set. The input data is checked and then system allows the triggering of the experiment. The measurement takes about 2 minutes to complete. The data is displayed in the user screen along with the plot as shown in Figure 5.

### 4. Uploading laboratory report:

After the experiment has been performed successfully by the learner, the experimental data is saved and analyzed by the student for the preparation of the laboratory report. The report contains detail from the experiment, setup, data and the results for parameter values etc. The online laboratory report submission is menu prompted. After the preliminary section of the laboratory report is submitted, the user has to enter the final results (parameter values) obtained from the analyses. These parameter values are then compared with the pre-set (default) parameter values generated by the instructor. Depending on the off-set (in percentage), the student is given marks and grading is done accordingly. The instructor sets some tolerance for evaluation purposes and the system compares the student's uploaded value with the default value. For example, for a tolerance of 15% only, 2 marks will be deducted from the full marks for the parameter. The instructor may specify up to 3 tolerances levels and the corresponding marks that are to be deducted.

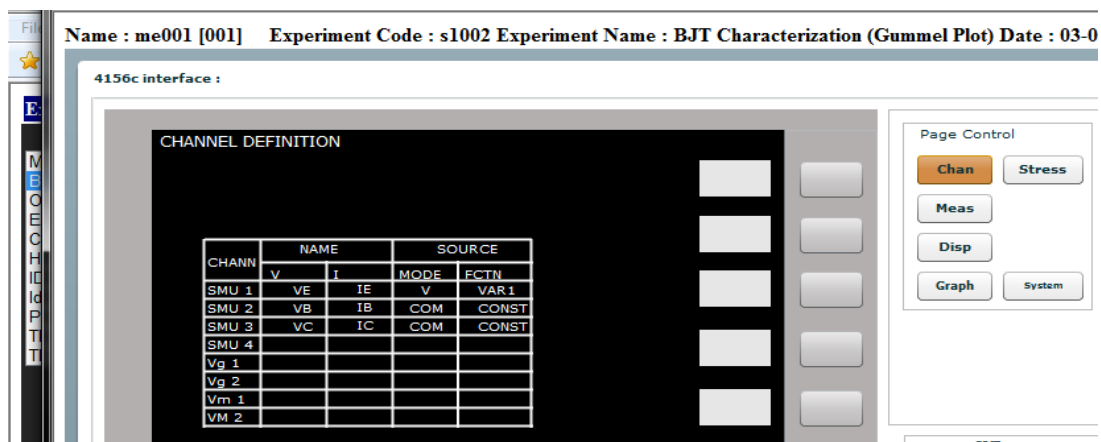


Figure 4(a). Channel definition setup in 4156C in the user interface

## 5. Viva voce:

After the students perform the experiment and submit the online laboratory report, the students are evaluated for their understanding of the experiment through a viva-voce. Several viva questions (in dynamic mode) are set by the instructor for the experiments. The viva questions are either multiple choice questions (MCQ) or true/false type questions. When the students appear in the viva, a set of questions are displayed with the choice and the students have to answer in a specified time.

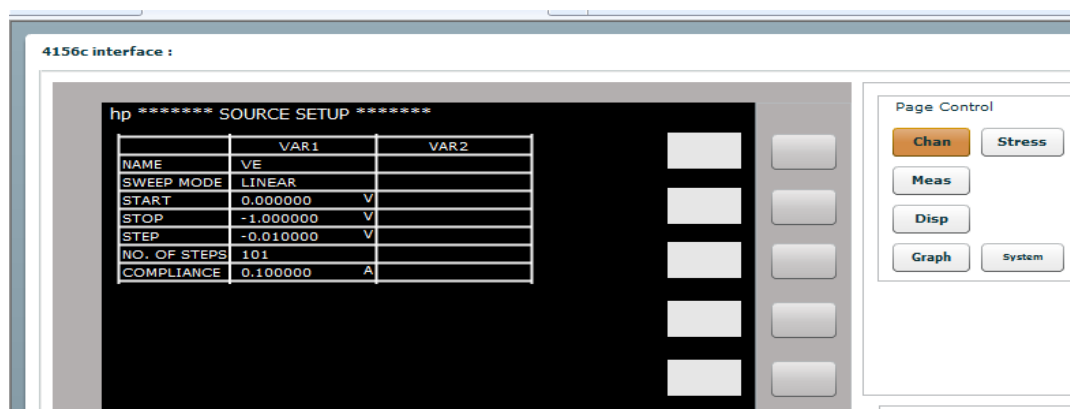


Figure 4(b). Source Setup in 4156C in the user interface

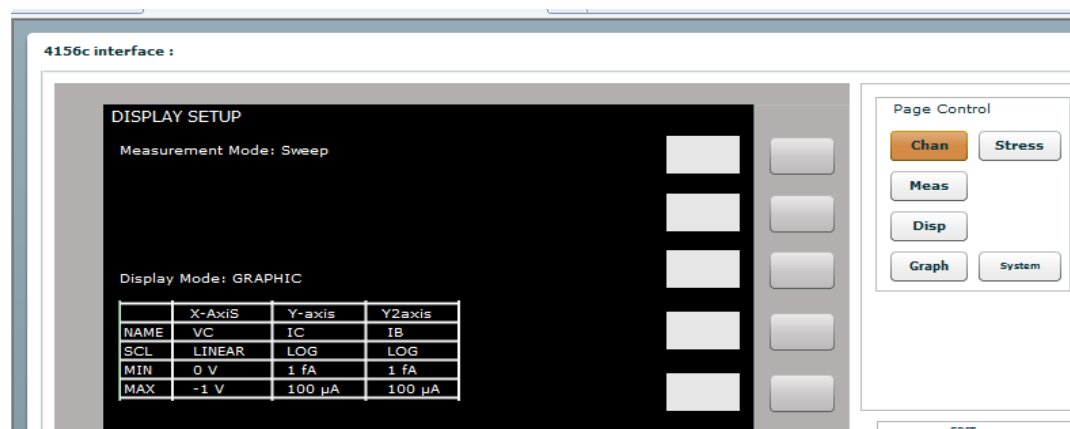


Figure 4(c). Display Setup in 4156C in the user interface



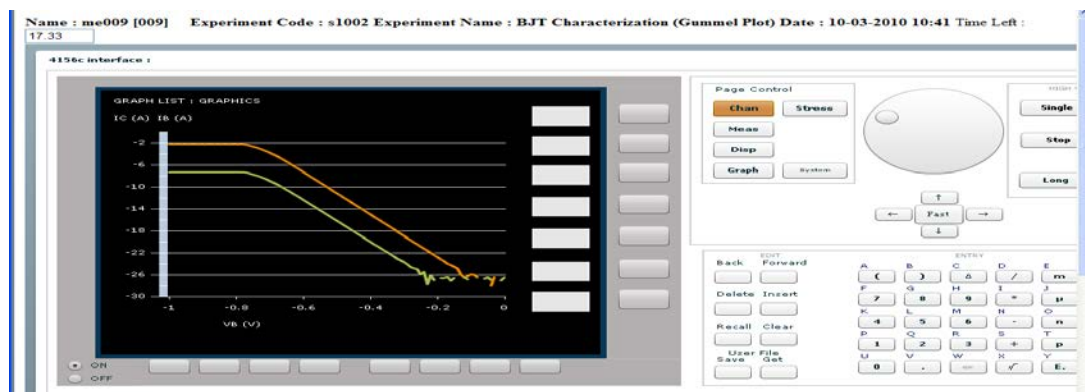


Figure 5. Gummel plot display on 4156C as seen in the user interface

## Evaluation of remote laboratory

Followings are some of the feedbacks and the results of the students and instructors activity profile (usage statistics) and opinion about NETLab as a learning tool. This study is based on a student population of about 40 for the period January-April 2010. All the students used the same website which has the functionality such as choice of experiment, experimental procedure, laboratory manual, pre-laboratory theory etc. Interactive session with teachers and students, as well as an inquiry/feedback, formed the basis for the evaluation. Evaluation procedures consist of several criteria, such as, Very Poor (VP), Poor (P), Fair (F), Good (G) and Very Good (VG) as measures. Student Feedback Form was used regularly which helped in receiving online feedback regarding their laboratory exercises. Laboratory Assessment Tools were used for evaluating the effectiveness of the laboratory experiments in terms of pedagogical benefits, resource utilization and cost effectiveness on a regular basis. Analytical Tools were provided in each experiment to help in performing various analyses on the data collected during experiment. Following typical parameters were used for evaluation:

- User Friendliness
- Freedom of Access
- Unmanned Assistance
- Assessment Support
- Experiment's Effectiveness in Learning
- Freedom of Creative Experimentation
- Ease of Modifying experiment
- Skill Developed
- Operational Control
- Availability of Analytical Tools
- Performance Effectiveness

Over a larger period of time it is observed that students prefer to perform experiments on weekdays rather on the weekends (see Figure 6). Five experiments mentioned in Section VIII use the same instruments and are similar in nature. Also the students performed all related and similar experiments in quick succession rather performing them in different sessions (see Figure 7). The results also show that a group of students performed the five experiments within a day.

As discussed in Section V, the NETLab LaMS contain an 'administrator section' for the teachers and administrator to keep control and track on the system (Logbook). The teachers can check the students' entry/exit records for information for future reference. The administrator can create new experiments and provide manuals, videos, images, animations etc. for each new experiment to update the laboratory module. Teachers are permitted to check the student's laboratory report and answer to the student questionnaire raised in the question box. The Logbook section also displays information about the student progress, the grade obtained and the time and period when the students performed the experiment.

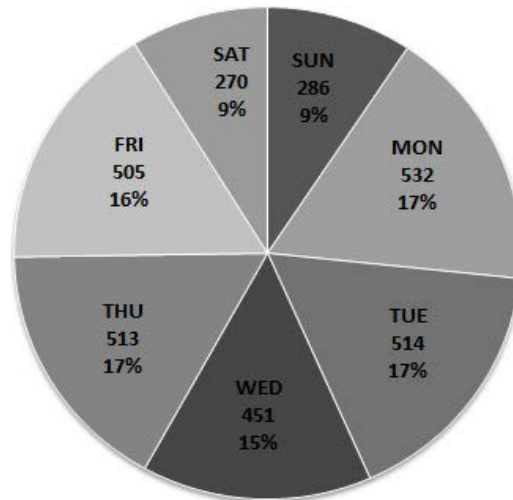


Figure 6. Percentage of users using the system during the days of week (2010)

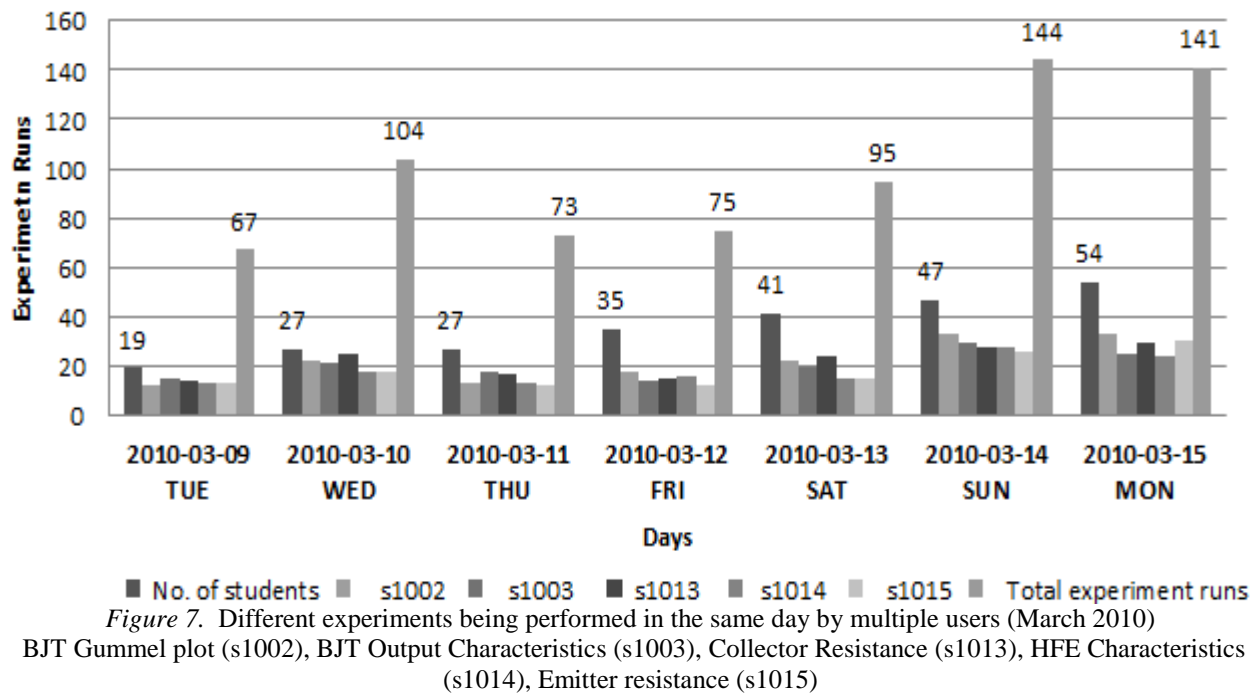


Figure 7. Different experiments being performed in the same day by multiple users (March 2010)

BJT Gummel plot (s1002), BJT Output Characteristics (s1003), Collector Resistance (s1013), HFE Characteristics (s1014), Emitter resistance (s1015)

## Conclusions

This paper addresses the issues involved in the design of experiments (with emphasis on electronics laboratory) and their implementation in remote laboratories employing internet technologies, both at the client and the server sides. Requirements related to security, quality of service, and operation and their implementation in remote laboratory management system are also discussed. Our experience with the three remote online laboratories described and particularly with the Microelectronics and VLSI Engineering Laboratory shaped our belief that an integrated design approach should facilitate the scalability of laboratory development and user management. The integrated design approach promotes student interaction with good infrastructure management that can ensure effective learning, better student performance, and achievement of the pedagogical goals.

This paper demonstrates the feasibility of designing remote laboratory systems for strong student interaction with remote equipment handling. The web-based, remote interactive laboratory environment allows remote students to access and utilize equipment located at a central equipment facility. The architectural design and the instructional strategies employed may be tailored to accommodate the special hardware and software requirements. This paper contributes to existing remote laboratory education and research by demonstrating the feasibility of designing the remote laboratory framework which employs secure, real time, interactive laboratories and incorporates effective online laboratory learning strategies, skill-building techniques including appropriate pedagogy, facilitation, and to impart knowledge to meet educational outcome.

## References

- Auer, M., Pester, A., Ursutiu, D., & Samoila, C. (2003, December). *Distributed virtual and remote labs in engineering*. Paper presented at the IEEE International Conference on Industrial Technology (ICIT), Maribor, Slovenia.
- Campbell, J. O., Bourne, R. J., Mosterman, P. J., & Brodersen, J. A. (2002). The effectiveness of learning simulators in electronic laboratories. *Journal of Engineering Education*, 91(1), 81-87. doi: 10.1002/j.2168-9830.2002.tb00675.x
- Fjeldly, T. A., & Shur, M. S. (Eds.) (2003). *Lab on the web—Running real electronics experiments via the Internet*. Hoboken, NJ: Wiley.
- Gomes, L., & Garcia-Zubia, J. (Eds.) (2007). *Advances on remote laboratories and e-learning experiences*. Bilbao, Spain: Univ. Deusto.
- Gomes, L., & Bogosyan, S. (2009). Current Trends in Remote Laboratories. *IEEE Transactions on Industrial Electronics*, 56(12), 4744-4756. doi: 10.1109/Tie.2009.2033293
- Guimaraes, E. G., Cardozo, E., Moraes, D. H., & Coelho, P. R. (2011). Design and Implementation Issues for Modern Remote Laboratories. *IEEE Transactions on Learning Technologies*, 4(2), 149-161. doi: 10.1109/TLT.2010.22
- Harward, V. J., Del Alamo, J. A., Lerman, S. R., Bailey, P. H., Carpenter, J., DeLong, K., . . . Zych, D. (2008). The iLab shared architecture a web services infrastructure to build communities of Internet accessible laboratories. *Proceedings of the IEEE, USA*, 96(6), 931-950. doi: 10.1109/Jproc.2008.921607
- Maiti, A. (2010a). NETLab: An Online Laboratory Management System. *International Journal of Online Engineering*, 6(2), 31-36. doi: 10.3991/ijoe.v6i2.1292
- Maiti, A. (2011a, April). *A Hybrid Algorithm for Time Scheduling in Remotely Triggered Online Laboratories*. Paper presented at the Global Engineering Education Conference (EDUCON), Amman, Jordan.
- Maiti, A., & Tripathy, B. K. (2011b, September). *An Improved Scheduling Scheme for the Management of Online Laboratories*. Paper presented at the Recent Advances in Intelligent Computational Systems (RAICS), India.
- Maiti, A. (2010b). Automatic evaluation of student's performance in online laboratories. *International Journal of Online Engineering*, 6. doi: 10.3991/ijoe.v6i5.1362
- Maiti, A., & Tripathy, B. K. (2011c). Applications of short message services in remotely triggered laboratories. *International Journal of Online Engineering*, 7(4), 20-25. doi: 10.3991/ijoe.v7i4.1809
- Maiti, A., & Maiti, A. (2010c). ADOBE Flash lite-based online laboratory for mobile phones. *International Journal of Interactive Mobile Technologies*, 4(4), 45-48. doi: 10.3991/ijim.v4i4.1440
- Reilly, R. (2008). Guest editorial virtual laboratories: Enhancing deep learning in model-based knowledge domains. *IEEE Transactions on Education*, 51(1), 1-1. doi: 10.1109/Te.2008.916030
- Sivakumar, S. C., Robertson, W., Artimy, M., & Aslam, N. (2005). A Web-Based Remote Interactive Laboratory for Internetworking Education. *IEEE Transactions on Education*, 48(4), 586-598. doi: 10.1109/TE.2005.858393
- Travis, J., & Kring, J. (2006). *LabVIEW for Everyone*. Upper Saddle River, NJ: Prentice-Hall.
- Trevelyan, J. P. (2004). *Lessons learned from 10 years' experience with remote laboratories*. Paper presented at the International Conference on Engineering Education and Research, V-TUO, Ostrava.
- Salih-Hassane, H., Saad, M., Ofosu, W. K., Mayaki, H. A., & Amadou, M. M. D. (2011). *Lab@Home: Remote laboratory evolution in the Cloud Computing Era*. Paper presented at the American Society for Engineering Education Annual Conference and Exposition 2011, Vancouver, Canada.

## Effects of Segmented-Animation in Projected Presentation Condition

**Ahmad Zamzuri Mohamad Ali**

Faculty of Art, Computing and Creative Industry, Universiti Pendidikan Sultan Idris, 35900, Tanjong Malim, Perak, Malaysia // zamzuri@fskik.upsi.edu.my

(Submitted March 06, 2012; Revised June 21, 2012; Accepted August 08, 2012)

### ABSTRACT

The aim of this study is to investigate the effects of segmented-animation and continuous-animation presentation in facilitating the learning of students, with different spatial visualization ability. An animation showing the cellular signal transmission process was developed for the purpose of this research. It was tested under a teacher-controlled projected presentation condition, which makes it different from many of the other previous similar studies. Pre-test and post-test experimental designs were used on two different groups. Spatial visualization ability was determined through a SVAT test. The data collected was statistically analyzed using Univariate Analysis of Covariance (ANCOVA). The results indicate that segmented-animation, which was presented with an LCD projector on a large white screen, was significantly more effective than continuous-animation in enhancing students' learning performance; specifically in memory recollection aspects. The results also show that segmented-animation presentation was beneficial for both low and high spatial visualization ability students, in conducting necessary cognitive processes, in order to develop more accurate mental model of the information depicted in the animation.

### Keywords

Continuous-animation, Mental model, Projected presentation, Segmented-animation, Teacher-controlled, Visualization

### Introduction

The utilization of animation is a key component of instructional courseware design nowadays. New advances in software technology have spurred the development of new means of conveying information in an animated form. Animation is defined as a series of rapidly changing computer screen displays that represent the illusion of movement (Mayer, 2001; Rieber & Hannafin, 1988). Specifically, instructional animation is defined as pictures in motion typically used to facilitate the instructional and learning (Lin, 2011). As such, animation has a potential role in supporting the visualization of a dynamic process, such as those not easily observable in real space and time scales, real processes that are practically impossible to realize in a learning situation, or a process that is not inherently visual (Betrancourt, 2005). Animation also plays a potential role in reducing the cognitive cost of mental simulation, thus saving cognitive resources for learning tasks; especially for novice learners or learners with poor imagery or spatial visualization capabilities (Betrancourt, 2005). In its best uses, animation presents information in a more interesting, easier to understand, and memorable way, than static media (Norton & Sprague, 2001; Rieber, 1990). However, even with these advantages and theoretical support, research findings related to the effectiveness of animation learning remain inconsistent (Lin & Dwyer, 2010; Ayres, Marcus, Chan & Qian, 2009; Ainsworth, 2008; Hegarty, 2004; Lin & Dwyer, 2004; Sperling, Seyedmonir, Aleksic & Meadows, 2003). The main reasons for inconsistencies may be attributed to design aspects (Tversky, Morrison & Betrancourt, 2002; Liu, Jones & Hemstreet, 1998) and learners' learning characteristics, such as their spatial visualization ability, learning style, motivation, prior knowledge, age, gender, and so on (Spanjers, van Gog & van Merriënboer, 2010; Mayer & Moreno, 2002; Chuang, 1999). There are dozens of models and theories on learning characteristics. However, this study only focuses on the spatial visualization aspects that are highly correlated with animation learning (van Oostendorp, Beijersbergen & Solaimani, 2008; Hegarty & Waller, 2005; Cornoldi & Vecche, 2003). It appears that current approaches to the design and use of animation can be ineffective due to the instructional designers' failure in addressing information processing challenges, posed by learners (Lowe, 2004). Therefore, approaches that lead to a reduction of these heavy cognitive processing needs are required especially to assist learners with poor imagery or spatial visualization ability (van Oostendorp, Beijersbergen & Solaimani, 2008).

### Animation and cognition

The theoretical framework of this study was based on Mayer's Cognitive Theory and Sweller's Cognitive Load Theory. Mayer's Cognitive Theory (Mayer, 2001) explained that information is processed in human memory through

two channels, namely the verbal channel and the visual channel, and through three cognitive processes. The first cognitive process involves the selection of verbal information to be processed in the verbal working memory and selecting the visual information to be processed in the visual working memory. The second cognitive process involves organizing the selected verbal information into a verbal mental model and organizing the selected visual information into a visual mental model. The final process involves integrating the verbal mental model and the visual mental model, developed with prior knowledge, to be stored in long-term memory. Successful learning occurs when learners are able to attend to important aspects of the presented material, mentally organize it into meaningful cognitive structure, and integrate it with relevant existing knowledge (Mayer & Moreno, 2003; 2002).

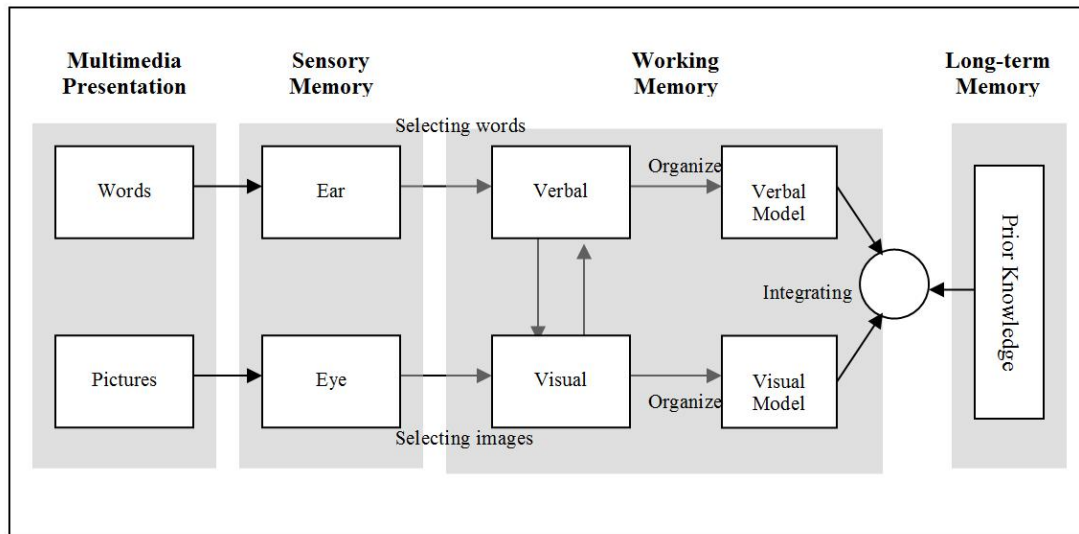


Figure 1. Mayer's model of memory (Source: Mayer, 2001)

Sweller's Cognitive Load Theory (Sweller, 2002; 1994) described learning structures in terms of an information processing system involving long-term memory, which effectively stores all the information gained on a more-or-less permanent basis, in schema form. Information may only be stored in long-term memory after first being attended to and processed by the working memory. However, the working memory is extremely limited in both capacity and duration. This limitation, under some conditions, may impede learning especially when learners have failed to develop an accurate mental model of the visual and verbal information in their working memory. The Cognitive Load Theory assumes that information should be structured, in order to eliminate any avoidable load on working memory, and enhance learning. The quality of instructional design will be raised if greater consideration is given to the role and limitation of the working memory.

Based on Mayer's and Sweller's theories, animation presentation that enables students to extract necessary information from it, in order to develop an accurate mental model of the information presented, is important for successful learning (Lowe, 1999). A continuous flow of changing information in animation may cause cognitive overload and hinder accurate mental model development (Betrancourt & Rebetez, 2007). This is due to the inability of learners to process new information, while simultaneously trying to remember and integrate important past information, thus creating an extraneous load as the working memory's resources are focused on dealing with demands of the presentation, rather than focusing on learning (Ayres, Marcus, Chan & Qian, 2009). However, there appears to be another view claiming that animation triggers passive information processing and may lead to overconfidence (Rozenblit & Keil, 2002). A study by Awan and Stevens (2006) revealed that confidence levels of understanding in an animation condition, was high in comparison with static media. They pointed out that higher confidence indicates lower mental effort, which leads to the perception that animation is easier to process. This situation can keep learners from doing relevant cognitive processing on their own (Schnotz & Rasch, 2005), which may influence the effectiveness of animation learning (van Oostendorp, Beijersbergen & Solaimani, 2008). Therefore, the need for empirical research to find design solutions that address these issues is important.

## **Segmented-animation**

Due to the speed and visual complexity of animation, learners may be confused and overwhelmed, throughout the animation viewing, as they are unable to properly process all incoming information before it disappears (Ayres, Marcus, Chan & Qian, 2009; van Oostendorp, Beijersbergen & Solaimani, 2008; Lowe, 2004; Weir & Heeps, 2003; Lowe, 1999). Therefore, designs that do not provide sufficient time for learners to focus their attention on the information being presented could be among the reasons for failure of animation to assist in learning (Torres & Dwyer, 1991). Without appropriate time allocation, learners may fail to develop new or adapt to existing schemas effectively (Garhart & Hannafin, 1986). It has been argued that the more time learners spend interacting with instructional material, the better they will be able to register the information in their long-term memory structure (Slater & Dwyer, 1996).

Segmented-animation, with features that allow learners to control the segment viewing rather than passively viewing the whole animation continuously, could be a design solution that addresses this limitation (Ahmad Zamzuri, 2007; Moreno, 2007; Mayer & Moreno, 2003; Mayer & Chandler, 2001). In segmented-animation design, the whole animation is broken into meaningful segments, including pauses or time breaks after each segment, and learner-control features to move from one segment another. Thus, pauses or time breaks between segments and learner-control features will allow learners to rehearse, in order to extract the necessary information from one segment before moving on to the next. In addition, during these pauses, learners can analyze the visual spatial structure of the content on the screen, which is something that can be difficult to do when a display continuously changes (Lowe, 2004).

Research findings have also revealed that segmented-animation; with learner-control features, will allow appropriate exposure duration of the animation that will help learners to interpret and understand the animation better (Ahmad Zamzuri, 2007; Aminordin, Ng & Fong, 2004; Fong, 2001). However, these studies were conducted in a fully learner-controlled self-paced learning condition. The question arises, would the same result would be obtained if the segmented-animation is in a teacher-controlled projected presentation learning condition, as moving from one segment to another is decided by the teacher and not on learners' preferences. Surprisingly, the study by Spanjers, Wouters, van Gog, and van Merriënboer, (2011) revealed that students learned more efficiently in segmented-animation than in continuous-animation, though without learner-control. In their study, the segmented-animation demonstrated and explained how probability calculation problems dealing with complex events need to be solved, automatically paused for 2 seconds before continuing. This study shows that segmentation still has positive effects on learning, even if it was not learner controllable (Spanjers, van Gog & van Merriënboer, 2010).

## **Spatial visualization ability**

Spatial visualization ability is a subset of spatial ability (Maizam, Black & Gray, 2002), which is the ability to understand and communicate using visual images (Casey & Wolf, 1989). In showing dynamic processes, the application of dynamic display techniques, such as animation, has a potential role in enhancing learners' visualization ability (Wiebe, 1991). Therefore, this ability seems to be relevant to the construction of an adequate mental model of the process depicted in the animation (van Oostendorp, Beijersbergen & Solaimani, 2008). It might be more effective than static media, especially for low spatial visualization learners, because they have less ability to mentally animate static diagrams (Boucheix & Schneider, 2009). The research findings also revealed that high spatial visualization ability learners primarily benefit from animation (Mayer & Sims, 1994, van Oostendorp, Beijersbergen & Solaimani, 2008, Wender & Muehlboeck, 2003). These research findings suggest that low spatial visualization ability students need more support in order to learn from animation better (van Oostendorp, Beijersbergen & Solaimani, 2008). Therefore, this study also investigated the effects of segmented-animation presentation, in assisting learning of students with different spatial visualization ability.

## **Method**

### **Research objective**

Based on the literature overview above, the main objective of this research is to study the effects of segmented-animation and continuous-animation on the learning achievements of students with high and low spatial visualization

ability. This study focuses on a teacher-controlled projected presentation condition. In detail, the predictions derived from the literature overview, are as follows:

- The teacher-controlled segmented-animation in a projected presentation condition has significant positive effects on a students' learning achievement compared to continuous-animation.
- The teacher-controlled segmented-animation in a projected presentation condition has significant positive effects on learning achievement of students with low spatial visualization ability compared to continuous-animation.
- The teacher-controlled Segmented-animation in projected presentation condition has significant positive effect on learning achievement of students with high spatial visualization ability compared to continuous-animation.

## Teaching material

An animation presentation with two different teacher-control strategies was developed. The animation, which shows a cellular signal transmission process, was embedded into the presentation's slide interface (Figure 2). The content of the animation are not easily observable in real space and time scales, practically impossible to realize in a learning situation, and is not inherently visual, which makes it suitable to be presented in an animation form for effective learning as pointed out by Betrancourt (2005). Furthermore, students need to visualize and imagine the content in actual three dimensional environments to understand the process depicted in the animation better. Therefore, animation appears to be a good method in presenting this content, since it has potential role in enhancing learners' visualization ability (Wiebe, 1991) and assists learning of low spatial visualization ability learners who have less ability to mentally animate static diagrams (Boucheix & Schneider, 2009).

Rationale of the interface design and content presentation strategy in reducing cognitive load has been detailed in Ahmad Zamzuri (2007; 2008). The appropriate content presentation design strategy to reduce any intrinsic load, and interface design to reduce any extraneous load, is important to avoid factors that might interfere with the schema acquisition by students throughout the study (Sweller, van Merriënboer & Paas, 1998). There was a study conducted using the earlier version of this courseware, but it was in a learner-controlled self-paced learning condition. In that study, students had full control over viewing of the animation based on their own preferences. Furthermore, the animation was displayed on each individual's personal computer screen. The study revealed that segmented-animation in a learner-control condition, was better than continuous-animation in enhancing students' learning achievement (Ahmad Zamzuri, 2007). However, this study will focus on a teacher-controlled projected presentation condition, where the pause duration between segments was determined by the teacher and the presentation was displayed on a big white screen.

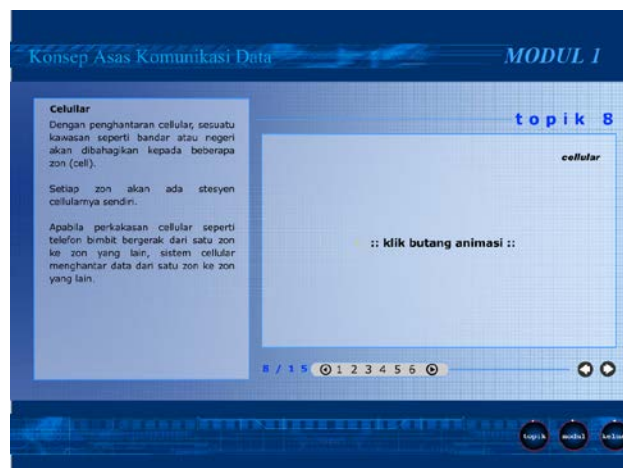


Figure 2. Presentation slide interface

The first animation presentation strategy was segmented-animation (Figure 3), which consisted of an animation that was broken into segments that the teacher could display in sequence using the 'next' and 'previous' control buttons. There appears to be no clear guidelines or design principles in the literature in determining the optimum segment length. However, the two options available, based on the researchers practice, were:

- The length is based on theories with regards to cognitive functioning.



- The length is based on the content expert's view, in order to determine the meaningful segments (Spanjers, van Gog & van Merriënboer, 2010).

This study adopted both options. The animation was broken into meaningful segments based on the expert's view. The segments were kept short, which was both necessary and helpful for novice learners, since all of the information was new for them (Spanjers, van Gog & van Merriënboer, 2010).

The complete animation display was approximately 27 seconds in length, which was then broken into 6 meaningful segments. The first segment showed zoning of location, the second showed placing of tower in each zone, the third and fourth segments showed communication within the same zone and the fifth and sixth segments showed communication between the different zones. Each segment displayed for approximately 4 to 5 seconds. The pause duration between segments could be decided by the teacher.

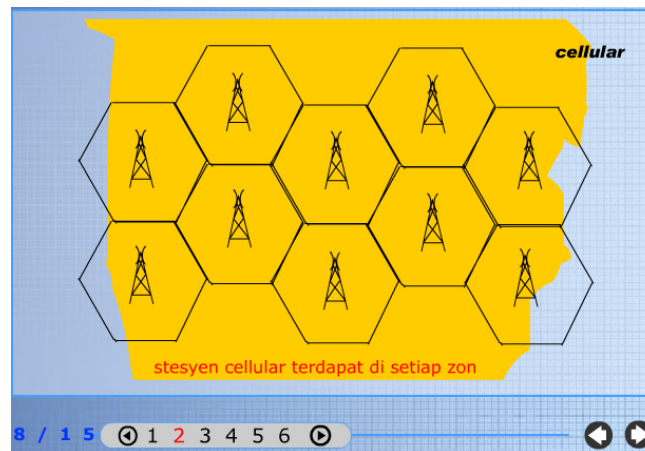


Figure 3. Segmented-animation strategy

The second animation presentation strategy was continuous-animation (Figure 4), which consisted of the same animation, but displayed continuously when the teacher clicked the 'start' button.

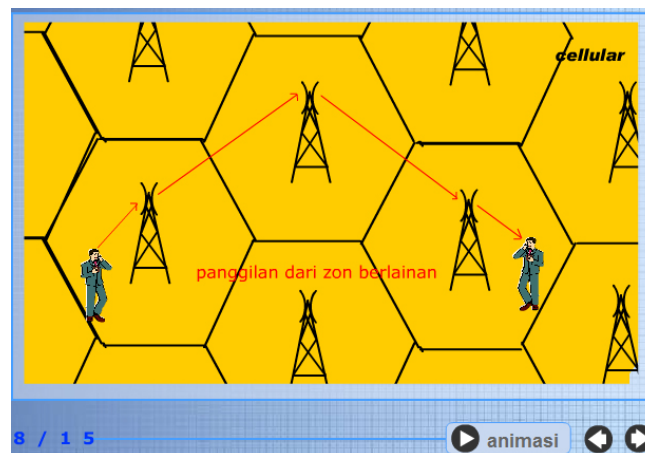


Figure 4. Continuous-animation strategy

## Procedure

This experimental study investigated the effects of two different animation presentation strategies on learning achievements of students with high and low spatial visualization ability. Independent variables were the segmented-animation and continuous-animation presentations, whilst the dependent variable was the post-test. Meanwhile, the moderator variables were the students' spatial visualization ability, whether they were high or low. The research



participants were 124 university students from two different classes, whose ages ranged from 22 to 26, and were enrolled in a Diploma in Education course. Basically, they have low prior knowledge on the content presented.

The study was conducted separately for both groups. The instructional material was presented to students with an LCD projector on a big white screen. They were given 60 seconds to read a projected text explanation, before viewing the animation. In the segmented-animation group, approximately 50 seconds pause duration, between segments, was determined by the teacher. This pause duration of 50 seconds was deemed suitable based on the teacher's preference, derived from his experience, expertise of the content, and knowledge of the students' ability. In the continuous-animation group, the teacher displayed the animation immediately after the time allocated for reading the text explanation. This process was repeated continuously three times in an identical manner for both groups. The purpose of replay was to ensure that students were provided sufficient time to extract necessary information from the animation. The number of replay was also determined by the teacher. The same teacher conducted the presentation for both groups. No verbal explanations from the teacher were given throughout the study as well.

### Test instrument

Pre-tests and post-tests were used on the two groups that viewed the two different animation presentation strategies respectively. Pre-tests and post-tests were written essay tests that only measured students' memory recall ability. The tests required students to draw and explain exactly the information depicted in the animation. Grades given were based on the accuracy and similarity of their answers to the animation presented. The pre-test was conducted before the learning process and the post-test was conducted immediately after the learning process. Both tests were similar. Two-way and one-way ANCOVA was employed to statistically analyze the collected data.

ASVAT test instrument, developed by Maizam (2000), was used to measure the students' visualization ability. This instrument consists of 28 items; 10 on the first task, which was a cube construction task, 11 on the second task, which was an engineering drawing, and 7 on the third task, which was a mental rotation task. The instrument had a Kuder-Richardson 20 (KR20) coefficient of 0.70 and a concurrent validity of 0.74 with the Vandenberg Mental Rotation test. The KR20 coefficients for task one, two, and three were 0.55, 0.43, and 0.53, respectively (Maizam, Black & Gray, 2002). The SVAT instrument has also been used in other studies (e.g. Ahmad Rizal, 2009) with acceptable reliability scores. The SVAT test was conducted 1 week before the commencement of the study.

### Results

A two-way ANCOVA was conducted in order to examine whether there was any significant difference in the learning performance of students in the segmented-animation condition and the continuous-animation condition. Independent variables were animation presentation strategy. The dependent variable was post-test scores, administrated following completion of the learning process. Scores of pre-test administrated prior to the commencement of the learning process were used as a covariate.

*Table 1. Tests between-subjects effects*

Source	<i>Df</i>	<i>Ms</i>	<i>F</i>	<i>P</i>	<i>Partial Eta Squared</i>
Pre test	1	1548.08	5.43	.021	.04
Mode	1	4497.78	15.77	.000	.12
Visualization	1	443.47	1.56	.215	.01
Mode * Visualization	1	276.47	.97	.327	.01
Error	119	285.17			

R squared = .153 (Adjusted R Squared = .124)

Preliminary checks were conducted to ensure that there was no violation of the assumptions of normality, linearity, homogeneity of variances, homogeneity of regression slopes, and reliable measurement of the covariate. After adjusting for pre-test scores, the interaction effect between animation presentation strategy and spatial visualization ability was not significant.  $F(1,119) = .97$ ,  $p > 0.05$ , with small effect size according Cohen's 1988 guidelines (partial eta squared = 0.01) as shown in Table 1. The plotted adjusted mean for the post test, split for high and low visualization ability and for the two animations presentation strategy clearly indicates no interaction occurs between

the two independent variables (Figure 5). For both high and low visualization ability, post-test scores were lowest in continuous-animation strategy and highest in segmented-animation strategy. This clearly suggests that high spatial visualization ability learners and low spatial visualization ability learners respond similarly in both animation presentation strategies.

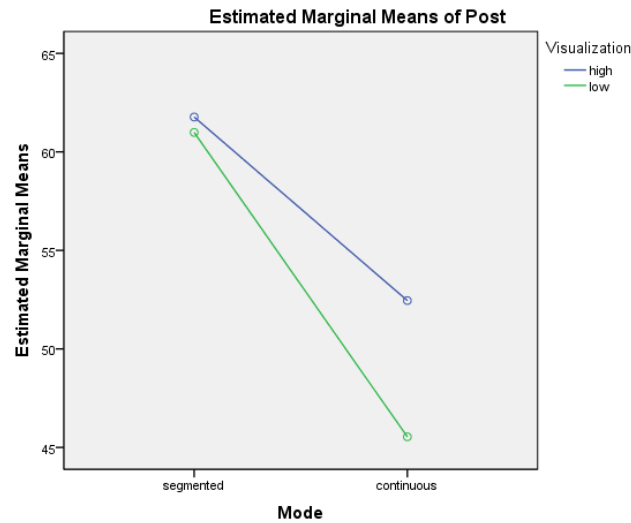


Figure 5. Interaction plot showing two simple main effects

The main effects for animation presentation strategy was statistically significant  $F(1,119) = 15.77$ ,  $p < 0.05$ , with medium effect size according Cohen's 1988 guidelines (partial eta squared = 0.12) as shown in Table 1. However, the main effects for spatial visualization ability was statistically not significant  $F(1,119) = 1.56$ ,  $p > 0.05$ , with low effect size (partial eta squared = 0.01) as shown in Table 1. This means that high spatial visualization ability and low spatial visualization ability learners do not differ in terms of their post-test scores, but there is a difference in scores for learners in segmented-animation and continuous-animation. The adjusted mean scores indicate that students in the segmented-animation condition obtained a better mean score ( $M=61.38$ ) than those students in the continuous-animation condition ( $M = 48.99$ ) as shown in Table 2. The results obtained indicate that overall, segmented-animation appeared to facilitate learning better than continuous-animation and therefore the first hypothesis is accepted.

Table 2. Descriptive statistics

Mode	Visualization	N	Unadjusted		Adjusted	
			M	SD	M	SE
segmented	High	25	60.80	17.78	61.77	3.40
	Low	39	60.77	16.60	60.99	2.71
	<b>Total</b>	64	60.78	16.93	61.38	2.18
continuous	High	34	53.53	19.68	52.45	2.93
	Low	26	45.38	13.63	45.54	3.31
	<b>Total</b>	60	50.00	17.66	48.99	2.21

Covariates appearing in the model are evaluated at the following values: Pre-test = 15.02

To explore the relation further, analysis was conducted by splitting the spatial visualization ability group and look at the effect of animation presentation strategy on post-test scores separately for low and high spatial visualization ability learners. A one-way ANCOVA was conducted in order to compare the effectiveness of two different animation presentation strategies. Preliminary checks were conducted to ensure that there was no violation of the assumptions of normality, linearity, homogeneity of variances, homogeneity of regression slopes, and reliable measurement of the covariate. After adjusting for pre-test scores, there was significant difference on post-test achievement of low spatial visualization ability learners between segmented-animation and continuous-animation condition,  $F(1,62) = 15.39$ ,  $p < 0.05$ , with large effect size (partial eta squared = 0.20) as shown in Table 3. The adjusted mean scores indicate that low spatial visualization ability students in the segmented-animation condition obtained a better mean score ( $M = 60.78$ ) than those low spatial visualization ability students in the continuous-

animation condition ( $M = 45.37$ ) as shown in Table 4. The results indicate that, segmented-animation appeared to facilitate low spatial visualization ability students' learning better than continuous-animation and therefore the second hypothesis is accepted.

Table 3. Tests between-subjects effects for low spatial visualization ability learners

Source	<i>Df</i>	<i>Ms</i>	<i>F</i>	<i>P</i>	<i>Partial Eta Squared</i>
Pre test	1	184.25	.77	.385	.01
Mode	1	3708.00	15.39	.000	.20
Error	62	240.95			

R squared = .206 (Adjusted R Squared = .180)

Table 4. Descriptive statistics for low spatial visualization ability learners

Mode	<i>N</i>	<i>Unadjusted</i>		<i>Adjusted</i>	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SE</i>
Segmented	39	60.77	16.60	60.78	2.49
Continuous	26	45.38	13.63	45.37	3.04

Covariates appearing in the model are evaluated at the following values: Pre-test = 14.51

Meanwhile, there was significant difference on post-test achievement of high spatial visualization ability learners between segmented-animation and continuous-animation condition,  $F(1,56) = 4.21$ ,  $p < 0.05$ , with medium effect size (partial eta squared = 0.07) as shown in Table 5. The adjusted mean scores indicate that high spatial visualization ability students in the segmented-animation condition obtained a better mean score ( $M = 62.50$ ) than those high spatial visualization ability students in the continuous-animation condition ( $M = 52.28$ ) as shown in Table 6. The results indicate that, segmented-animation appeared to facilitate high spatial visualization ability students' learning better than continuous-animation and therefore the third hypothesis is also accepted.

Table 5. Tests between-subjects effects for high spatial visualization ability learners

Source	<i>Df</i>	<i>Ms</i>	<i>F</i>	<i>P</i>	<i>Partial Eta Squared</i>
Pre test	1	1702.29	5.11	.028	.08
Mode	1	1401.41	4.21	.045	.07
Error	56	333.182			

R squared = .117 (Adjusted R Squared = .085)

Table 6. Descriptive statistics for high spatial visualization ability learners

Mode	<i>N</i>	<i>Unadjusted</i>		<i>Adjusted</i>	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SE</i>
Segmented	25	60.80	17.78	62.50	3.72
Continuous	34	53.63	19.68	52.28	3.18

Covariates appearing in the model are evaluated at the following values: Pre-test = 15.59

## Discussion

As pointed out earlier, there is much research concerning animation, and much theoretical support exists in the field of effective learning. However, the characteristics of animation means that changes over time may cause cognitive overload throughout the learning process that could hinder effective cognition especially for students with low spatial visualization ability (van Oostendorp, Beijersbergen & Solaimani, 2008). Therefore, segmentation was proposed as a potential solution in facilitating students' cognition in animation learning (Ahmad Zamzuri, 2007; Moreno, 2007; Mayer & Moreno, 2003; Mayer & Chandler, 2001). The findings from this research are in line with that segmentation proposal. Overall, the findings revealed that students in the segmented-animation condition obtained better mean scores than students in the continuous-animation condition. Results were consistent with many other studies on segmentation; in learner-control (Ahmad Zamzuri, 2007; Aminordin, Ng & Fong, 2004; Fong, 2001) and in program-control (Spanjers, Wouters, van Gog & van Merriënboer, 2011). Therefore, the results of this study extended the effectiveness of segmentation to include teacher-controlled projected presentation mode too.

It was noted that even though the pauses between segments does increase the total available learning time, this does not necessarily increase learning ability. A study by Hasler, Kersten, and Sweller (2007) and Ahmad Zamzuri (2007) showed that even if equal time was allocated for all conditions, segmented-animation still promoted a better learning outcome. In addition, as pointed out by Spanjers, van Gog, and van Merriënboer (2010), providing additional time at the end of an entire, continuous dynamic visualization does not necessarily enhance learning, because this would do nothing to alleviate the high load at certain points in time. Furthermore, sufficient time was allocated for the learners, by repeating the animation presentation three times, to enable them to extract necessary information and understand the information depicted in the animation better.

The results also indicate that segmented-animation was beneficial for students with low spatial visualization ability; even if it was not learner controllable and in the projected presentation condition. This finding further strengthens the assumption that animation, which is broken into meaningful segments, including pauses or time breaks between segments, will allow appropriate time for students with low spatial visualization ability to extract the necessary information in order to develop a more accurate mental model of the information depicted in the animation. With an accurate mental model, formation of meaningful schemas to be registered in long-term memory is possible; even for students with low spatial visualization ability. In addition, during the pauses, learners can analyze the visual spatial structure of the animation at certain meaningful points. Apparently, pauses play the role of a static display at certain meaningful points of an animation. Therefore, this situation could overcome the passive cognitive processing phenomenon that leads to overconfidence and negative animation learning outcomes, as pointed out by Awan and Stevens (2006). Questions might arise on the implication of pause duration on the students' learning outcome. In this study, the pause duration was determined by the teacher, based on his experience, expertise of the content, and knowledge of students' learning ability. This may vary for different content and different students' learning ability. Therefore, further investigation looking into the effects of different pause durations or time breaks between segments is important. The study by Spanjers, Wouters, van Gog, and van Merriënboer (2011) revealed that even a pause duration of as little as 2 seconds, promoted better learning in segmented-animation condition, in comparison with other conditions. However, the outcome of self-paced learning and projected presentation learning may differ. In the self-paced learning condition, students still have freedom in deciding the repetition process. Whereas, in a teacher-controlled projected presentation condition, the decision is solely the teacher's.

Students with high spatial visualization ability seem also benefited from the segmented-animation strategy. This finding is not in line with the assumption that high spatial visualization ability learners primarily benefit from the animation learning (Mayer & Sims, 1994; van Oostendorp, Beijersbergen & Solaimani, 2008, Wender & Muehlboeck, 2003). Even though, high spatial visualization ability learners in continuous-animation strategy obtained better mean score than low spatial visualization ability learners, high and low spatial visualization ability learners in segmented-animation strategy far out performed them in mean score terms. It is surprising to notice that high and low spatial visualization ability learners in segmented-animation strategy obtained almost similar mean scores. There might be many possibilities for this, which includes students' prior knowledge. Higher prior knowledge learners have to invest less mental effort into learning a given topic. Consequently, they have more cognitive capacity left for trying to understand a displayed motion concerning the topic on a very detailed level in comparison with lower prior knowledge learners (Höffler & Leutner, 2007). Learning is a process of receiving, processing, coding, storing and retrieving information from memory structure. As pointed out in introduction section, human memory structure is divided into three processing system namely sensory memory, working memory and long term memory. Sweller (2002) described learning structures in terms of an information processing system, which effectively stores all the information gained in long term memory on a more-or-less permanent basis in schematic form. Schemas are memory structures that allow learners to deal with a large number of information blocks as if they are a single block. New information entering working memory must be integrated with preexisting schemas in long term memory for adequate mental model development in working memory. For this to take place, relevant schemas in long term memory must be activated and decoded into working memory which is also defined as schema acquisition. Information may only be stored in long term memory after first being attended to and processed by the working memory. However, the working memory is extremely limited in both capacity and duration. Consequently, not all the information entered the working memory structure will be registered in long term memory in an adequate schema form. This limitation, under some conditions may impede learning, especially novices or low prior knowledge learners who are unable to recognize relevant preexisting schemas to solve problem. Based on this arguments, this research finding clearly indicates, high spatial visualization ability does not have any impact on novices or students with low prior knowledge. However, studies are needed to further affirm this assumption,

specifically on the correlation between prior knowledge and spatial visualization ability in segmented-animation learning.

Beside students' prior knowledge, content of the animation, complexity of the animation display, animation display speeds, animation length also might have influence. It is not an easy task to conclude a one-size-fits-all design strategy that would addresses all these influencing factors. Thus, findings of this study suggest segmentation as a potential design solution for effective animation learning for both low and high spatial visualization ability learners specifically in projected presentation condition.

## Conclusions

Development of computer animations is relatively time-consuming and costly. Therefore, to ensure the maximum effectiveness of animation learning, instructional designers should ground their designs based on current research findings and theories. They should not base their designs on their own preferences, which may or may not work well. Thus, the findings from this research suggest that segmented-animation presentation has an advantage in promoting better animation learning. Segmented-animation actually plays the role of both dynamic and static displays. During the segment viewing, animation assists learners into the construction of adequate mental models in their working memory. With an adequate mental model, the formation of more meaningful schemas to be registered in long-term memory is possible. Meanwhile, during the pauses, the static display promotes learners to complete relevant cognitive processes on their own, by integrating with previously developed schemas. Therefore, the combination of passive and active cognitions promoted by segmented-animation may play a potential role for successful learning, specifically for novices or low prior knowledge learners.

As pointed out in the discussion section, further study needs to be conducted on various segment lengths, display speeds, animation length, pause durations or time breaks between segments, and correlation between prior knowledge and spatial visualization ability to further strengthen this promising instructional animation design strategy. Since this study was limited to only measuring memory recall ability, further study on measuring higher cognitive ability such as analysis and synthesis ability as pointed in Bloom's taxonomy also seems important.

## Acknowledgements

The author wishes to acknowledge the support of the Research and Management Centre, Universiti Pendidikan Sultan Idris, who awarded a research grant for this study.

## References

- Ahmad Rizal M. (2009). Keberkesanan animasi grafik dalam kalangan pelajar berbeza gaya kognitif dan kebolehan visualisasi spatial di politeknik (*Effects of animation on polytechnic students with different cognitive styles and spatial visualization ability*). (Unpublished doctoral thesis). Universiti Teknologi Malaysia, Johor, Malaysia.
- Ahmad Zamzuri M. A. (2007). Kesan koswer strategi animasi pelbagai kawalan pengguna terhadap prestasi kognitif dan ketekalan pelajar (*Effects of various user-controlled animation strategy courseware on the students Cognitive and Retention Skills*). (Unpublished doctoral thesis). Universiti Sains Malaysia, Pulau Pinang, Malaysia.
- Ahmad Zamzuri M. A. (2008). Effective instructional courseware design to improve students' cognitive skills: A practical guide for educators as multimedia authors. In M. I. Rozhan, I. Issham, T. Zaidatun, S. F. Fong, & M. Balakrishnan (Eds.), *Proceedings of the 2<sup>nd</sup> International Malaysian Educational Technology Convention* (pp. 245-252). Kuala Lumpur, Malaysia: Malaysian Educational Technology Association.
- Ainsworth, S. (2008). How do animations influence learning? In D. H. Robinson & G. Schraw (Eds.), *Current perspectives on cognition, learning, and instruction: Recent innovations in educational technology that facilitate student learning* (pp. 37-67). Charlotte, NC: Information Age Publishing.
- Aminordin C. L., Ng, K.T., & Fong, S.F. (2004). Kesan paparan grafik dalam pembelajaran fenomena cuaca (*Effects of graphics in learning weather phenomena*). In A. Hanafi & M. I. Rozhan, *Proceedings of 17<sup>th</sup> Educational Technology Convention, Malaysia* (pp. 249-255). Kuala Lumpur, Malaysia: Persatuan Teknologi Pendidikan Malaysia.

- Awan, R. N., & Stevens, B. (2006). Static/animated diagrams and their effect on student's perceptions of conceptual understandings in Computer Aided Learning (CAL) environments. In McEwan, T., Gulliksen, J., & Benyen, D. (Eds), *People and computers XIX – The bigger picture* (pp. 381-389). London, UK: Springer-Verlag.
- Ayres, P., Marcus, N., Chan, C., & Qian, N. (2009). Learning hand manipulative tasks: When instructional animations are superior to equivalent static representations. *Computers in Human Behaviour*, 25(2), 348-353.
- Betrancourt, M. (2005). The animation and interactivity principles in multimedia learning. In Mayer, R.E. (Ed.), *The Cambridge handbook of multimedia learning* (pp. 287-296). New York, NY: Cambridge University Press.
- Betrancourt, M., & Rebetez, C. (2007, August). *Building dynamic mental models from animation: Effect of user control on exploration behaviours*. Paper presented at the 12<sup>th</sup> EARLI Biennial Conference on Research on Learning and Instruction, Budapest, Hungary.
- Boucheix, J., & Schneider, E. (2009). Static and animated presentation in learning dynamic mechanical systems. *Learning and Instruction* 19(2), 112-127.
- Casey, H. V. E., & Wolf, J. S. (1989). Developing visual literacy among academically able fifth-grade students. In D. E. Woodland, & L. F. Szul (1999), *Visualization ability, proofreading, and colour configurations of a computer screen-interaction and implications*. Retrieved from <http://www.osra.org/itlpj/woodlandszul.pdf>
- Chuang, Y. R. (1999). Teaching in a multimedia computer environment: A study of the effects of learning style, gender, and math achievement. *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning*. Retrieved from <http://imej.wfu.edu/articles/1999/1/10/>
- Cohen, J. W. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cornoldi, C., & Vecche, T. (2003). *Visuo-spatial working memory and individual differences: Essays in cognitive psychology*. Hove, UK: Psychology Press.
- Fong, S.F. (2001). Kesan animasi terhadap pembelajaran pengetahuan prosedur meiosis di kalangan pelajar pelbagai profil psikologi (*Effects of meiosis process procedure animation on learning performance of students with different psychological profiles*). (Unpublished doctoral thesis). Universiti Sains Malaysia, Pulau Pinang, Malaysia.
- Garhart, C., & Hannafin, M. (1986). The accuracy of cognitive monitoring during computer-based instruction. *Journal of Computer-Based Instruction*, 13(3), 88-93.
- Hasler, B. S., Kersten, B., & Sweller, J. (2007). Learner control, cognitive load, and instructional animation. *Applied Cognitive Psychology*, 21(6), 713-729.
- Hegarty, M. (2004). Dynamic visualization and learning: Getting to the difficult questions. *Learning and Instruction*, 14(3), 343-351.
- Hegarty, M., & Weller, D. A. (2005). Individual differences in spatial abilities. In P. Shah, P., & Myake, A. (Eds.). *The Cambridge handbook of visio spatial thinking* (pp. 121-169). New York, NY: Cambridge University Press.
- Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction*, 17(6), 722-738.
- Lin, C. L., & Dwyer, F. M. (2004). Effect of varied animated enhancement strategies in facilitating achievement of different educational objectives. *International Journal of Instructional Media*, 31(2), 185-199.
- Lin, H. (2011). Facilitating learning from animated instruction: Effectiveness of questions and feedback as attention-directing strategies. *Journal of Educational Technology and Society*, 14(2), 31-42.
- Lin, H., & Dwyer, F. M. (2010). The effect of static and animated visualization: A perspective of instructional effectiveness and efficiency. *Educational Technology Research and Development*, 58(2), 155-174.
- Liu, M., Jones, C., & Hemstreet, S. (1998). Interactive multimedia design and production processes. *Journal of Research on Computing in Education*, 30(3), 254-281.
- Lowe, R. K. (1999). Extracting information from an animation during complex visual learning. *European Journal of the Psychology of Education*, 14(2), 225-244.
- Lowe, R. K. (2004). Animation and learning: Value for money. In R. Atkinson, C. McBeath, D. Jonas-Dwyer, & R. Philips (Eds), *Beyond the comfort zone: Proceedings of the 21<sup>st</sup> ASCILITE Conference, Australia* (pp. 558-561). Retrieved from <http://www.ascilite.org.au/conferences/perth04/procs/lowe-r.html>
- Maizam A. (2000). *Spatial visualization ability and civil engineering problem solving*. (Unpublished doctoral thesis). University of Surrey, Guilford, UK.

- Maizam A., Black, T. R., & Gray, D. E. (2002). Effect of instruction on spatial visualization ability in civil engineering students. *International Educational Journal*, 3(1), 1-12.
- Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology*, 86(3), 389-401.
- Mayer, R. E., & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology*, 93(2), 390-397.
- Mayer, R. E. (2001). *Multimedia learning*. Cambridge, UK: Cambridge University Press.
- Mayer, R. E., & Moreno, R. (2002). Animation as an aid to multimedia learning. *Educational Psychology Review*, 14(1), 87-99.
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38(1), 43-52.
- Moreno, R. (2007). Optimising learning from animations by minimizing cognitive load: Cognitive and affective consequences of signalling and segmentation methods. *Applied Cognitive Psychology*, 21(6), 765-781.
- Norton, P., & Sprague, D. (2001). *Technology for teaching*. Needham Heights, MA: Allyn & Bacon.
- Rieber, L. P., & Hannanfin, M. J. (1988). Effects of textual and animated orienting activities and practice on learning from computer-based instruction. *Computer in Schools*, 5(1/2), 77-89.
- Rieber, L. P. (1990). Animation in computer based instruction. *Educational Technology Research and Development*, 38(1), 77-86.
- Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, 26(5), 521-562.
- Schnotz, W., & Rasch, T. (2005). Enabling, facilitating, and inhibiting effects of animations in multimedia learning: Why reduction of cognitive load can have negative results on learning. *Educational Technology Research & Development*, 53(3), 47-58.
- Slater, R.B., & Dwyer, F. (1996). The effect of varied interactive questioning strategies in complementing visualized instruction. *International Journal of Instructional Media*, 23(3), 273-280.
- Spanjers, I. A. E., van Gog, T., & van Merriënboer, J. J. G. (2010). A theoretical analysis of how segmentation of dynamic visualizations optimizes students' learning. *Educational Psychology Review*, 22(4), 411-423.
- Spanjers, I. A. E., Wouters, P., van Gog, T., & van Merriënboer, J.J.G. (2011). An expertise reversal effect of segmentation in learning from animations. *Computers in Human Behaviour*, 27(1), 46-52.
- Sperling, R. A., Seyedmonir, M., Aleksic, M., & Meadows, G. (2003). Animations as learning tools in authentic science materials. *International Journal of Instructional Media*, 30(2), 213-222.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295-312.
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251-296.
- Sweller, J. (2002). Visualization and instructional design. *Proceedings of the International Workshop on Dynamic Visualization and Learning, Germany* (pp. 1501-1510). Retrieved 24 May, 2010 from <http://www.iwm-knrc.de/workshops/visualization/sweller.pdf>
- Torres, J., & Dwyer, F. (1991). The effect of time in instructional effectiveness of varied enhancement strategies. *International Journal of Instructional Media*, 18, 26-31.
- Tversky, B., Morrison, J. B., & Betancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human Computer Studies*, 57, 1-16.
- van Oostendorp, H., Beijersbergen, M. J., & Solaimani, S. (2008). Conditions for learning from animations. *Proceedings of the 8<sup>th</sup> International Conference on Learning Science, The Netherlands* (pp. 438-445). Retrieved from <http://dl.acm.org/citation.cfm?id=1599871.1599926>
- Weir, G. R. S., & Heeps, S. (2003, August). *Getting the message across: Ten principles for web animation*. Paper presented at the 7<sup>th</sup> IASTED International Conference on Internet and Multimedia and Applications. Retrieved from <http://strathprints.strath.ac.uk/2513/1/strathprints002513.pdf>
- Wender, K. F., & Muehlboeck, J. S. (2003). Animated diagrams in teaching statistics. *Behaviour Research Methods, Instruments & Computers*, 35(2), 255-258.
- Wiebe, E. N. (1991). *A review of dynamic and static visual display techniques*. Retrieved Dec 28, 2004 from <http://www4.ncsu.edu/~wiebe/articles/ani-1991.pdf>

## The Framework of Viewing and Representing Skills Through Digital Text

Kay Yong Khoo\* and Daniel Churchill

<sup>1</sup>The University of Hong Kong, Pokfulam Road, Hong Kong // <sup>2</sup>The University of Hong Kong, Pokfulam Road, Hong Kong // kykhoo@yahoo.com // dchurch@hkucc.hku.hk

\*Corresponding author

(Submitted February 21, 2012; Revised August 01, 2012; Accepted August 30, 2012)

### ABSTRACT

This paper reports on case studies of four primary school children and their digital practices in Hong Kong. The study explores how the participating children view and represent through digital text in the context of their out-of-school technology use. Understanding how these practices extended into their English language classrooms was explicated based on the data that emerged from the study. The study results identify five aspects of emerging skills acquired by the children. These skills were not extended into the children's schoolwork – a reflection on the emphasis on mono-modal language learning in the class setting. A detailed examination of the individual skills led to a set of recommendations for curriculum review, suitable pedagogical strategies and classroom learning resources that English Language Educators may utilize to facilitate development of students' viewing and representing skills through digital text.

### Keywords

Digital text, Literacies, Viewing and representing skills, Multimodality

### The digital practices of primary school children

The meaning of literacy in the 21<sup>st</sup> century has changed markedly with emerging and now dominant technologies: that is, the move away from writing to the new digital text mode, and from the medium of the book to that of the monitor screen (Kress, 2010). It is therefore important to extend the notion of literacy beyond the traditional modes of listening, speaking, reading and writing to include digital text and communication.

Language curricula in primary schools around the world are being revised in countries such as the United Kingdom, Canada, Singapore and Australia to incorporate new forms of literacies (Department for Children schools and families UK, 2009; Ministry of Education Ontario, 2006; MOE Singapore, 2010; National Curriculum Board Australia, 2008). These literacies are proposed as prerequisites for the effective consumption of digital texts and communication through contemporary technologies (Kress & Jewitt, 2003; Unsworth, 2008).

Today's technologies allow children to engage with digital text on a regular basis outside school (Buckingham & Willett, 2006; Kimber & Claire, 2008; Prensky, 2001a). However, the extra-school skills that are acquired from interacting and communicating with digital text remain unexplored. Digital text incorporates the four macro-skills of listening, speaking, reading and writing, but requires additional skills including frequent use of visuals, dynamic information, and interaction.

Contemporary language curricula in schools need to incorporate emerging literacies (Jewitt, 2006; Kress, 2003; Martinec & Leeuwen, 2009). Language is broadly understood as the main meaning-making resource for constructing and consuming information (Unsworth, 2001). However, in Hong Kong, the government-mandated primary school English Language curriculum has not incorporated digital text as an integral aspect of language learning (The Curriculum Development Council, 2004). To incorporate digital text into the language curriculum, a viable set of recommendations is needed to address the skills required to engage with the emerging literacies.

There is an emerging gap between children's out-of-school digital practices and practices in the primary school English Language classroom in Hong Kong (which remain focused on linguistic learning). The literature suggests four areas of digital practices that impact the engagement on monitor screens:

- The ability to recognize and create elements in different modes (Jewitt, 2006; Unsworth, 2008);
- The ability to apply the affordances of modes in meaning making (Jewitt, 2006; Kress, 2003; Martinec & Leeuwen, 2009);



- The ability to link elements of information in different modes contextually in spatial or temporal layouts (Kress, 2010; Leeuwen, 2005); and
- The ability to navigate on screens (Leeuwen, 2005; Martinec & Leeuwen, 2009).

These four areas of digital practices are also evidenced among the children, who participate in various social networking sites regularly. Their semantic practices involve not just engagement with written text, but also elements of other modes (Connelly, 2008; Cope & Kalantzis, 2000; Unsworth, 2008). The escalation of the children's engagement on different sites has resulted in their quickly adapting to the different navigation potential of screens (Gee, 2003; Prensky, 2001b). They are also obtaining screen-based information with different layouts of different mode composition (Walsh et al., 2007).

Based on this concise review, we may conclude that there are skills to be derived from children's intuitive understandings arising from their natural interactions with digital text. The skills utilize but are additional to listening, speaking, reading and writing, and involve frequent use of visuals, dynamic information, and interaction through digital text. Such skills are categorized in the literature as viewing and representing (Kress, 2010; MOE Singapore, 2010).

Hong Kong is currently in a state of transition in which the primary school English Language curriculum has not yet incorporated changes in the new literacies. The purpose of this study is to investigate in detail aspects of participating children's viewing and representing skills and how they extended their viewing and representing skills into their formal schoolwork and independent school projects.

## Methodology

The study focused on four primary school-aged children and how the practices emerging from their out-of-school use of digital technology extended to formal school activities, whether initiated by teachers in formal schoolwork, or applied independently by the children in their independent schoolwork. The methodology of the study was qualitative (Gay et al., 2006; Merriam, 1988, 2009; Yin, 1994). Two research questions guided the data collection and analysis of this study:

- What digital literacy practices can be observed and identified from the participating children's out-of school technology use?
- How have these practices been extended to school activities?

Students were recruited using a purposeful sampling strategy at a tuition centre in Hong Kong. The centre operates primary school English Language tuition classes supplementary to the Hong Kong government school syllabus. It has implemented an e-learning system to support delivery of its program. Four children were selected according to three selection criteria determined by the researcher to ensure rich qualitative data. These criteria were: (i) participants regularly viewed or represented messages in blogs or social media tools like Facebook; (ii) used the Internet at least weekly and (iii) were students from primary-six level, the highest level of primary school. The research also investigated the children's digital-related schoolwork in their school.

Data was collected over a period of eight months. The study included observations, video recordings, interviews and personal documents (digital artifacts) study and was structured into two stages (see Figure 1). The study was explorative. The constant comparative method of data analysis was employed in this research. It involved collection of visual and narrative data and digital artifacts during the two stages. Stage One focused on developing a framework for viewing and representing the relevant skills of the participating children according to Mayer's Cognitive Theory of Multimedia learning (Mayer, 2005). Stage Two focused on understanding how the children extended these skills into their school activities. Understandings emerging from the first stage of the study facilitated data collection in the second stage. Findings from stage two of the study led to a viable set of recommendations for incorporating the skills required to engage with the emerging literacies into the language curriculum.

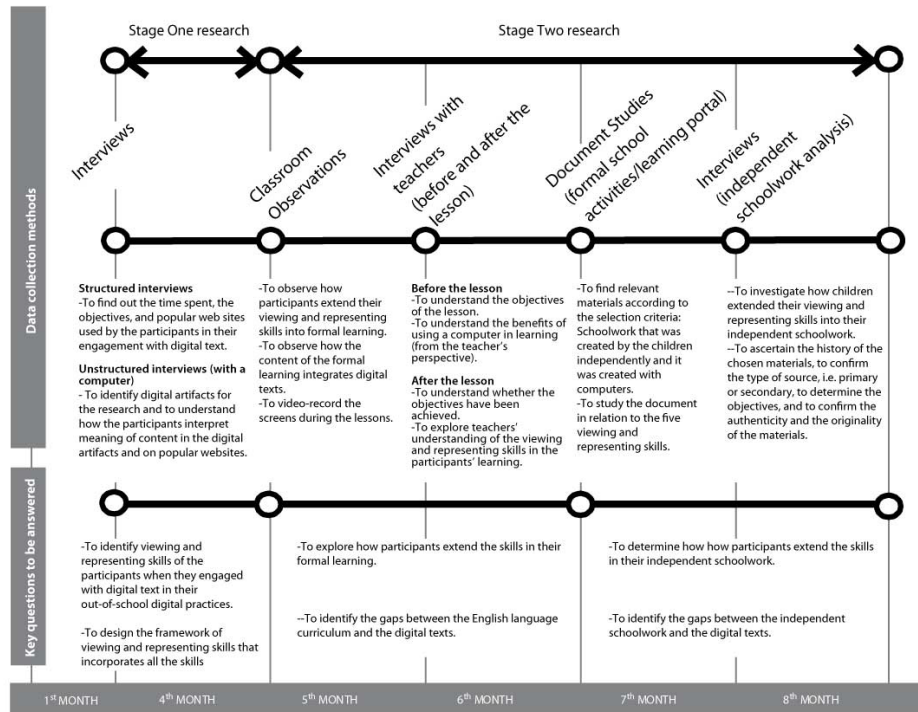


Figure 1. The Procedure of the research methodology

## Participants

Four cases were identified after discussing the selection criteria with the teachers. The participating children were from different schools and were enrolled in the same tuition centre.

Table 1. The details derived from the selection criteria for the four cases

Participant *	Hours/week Spent with Computers	Regular Viewing and Representing Activities	Gender/Level
Angel	16	Facebook.com	Female/ Primary 6
Winnie	4	myblog.yahoo.com	Female/ Primary 6
Nicole	3	Qooza.hk	Female/ Primary 6
Wong	4.5	Facebook.com	Female/ Primary 6

The study participants' profiles are presented in Table 1. Pseudonyms are used to safeguard their anonymity. The study was conducted after relevant ethical clearance was secured. Initially, a group of children were briefed about the study and roles of participants before being asked to consider participating. Those who agreed to participate were given a consent form to take home to their parents. Signed parental consent was required.

## Results: The digital skills developed by the participants

### Participant one: Angel, a multi-mode representing practitioner

Angel is an eleven-year-old girl from a middle-class Hong Kong Chinese family. Digital technology is an important part of her life. She claimed that her hobby was to design with graphic design software, such as Windows Paints and Photoshop. She also stated that she communicated regularly with her friends on Facebook; she read and replied to messages from her friends and created new messages in her social networking account. The researcher found that in her Facebook messages, the information was multi-mode - i.e. videos, images, verbal narration and writing. From day one of the study, Angel demonstrated a strong tendency to seek out information on the Internet, including

looking up the meaning of words with online dictionaries. The researcher asked Angel to post interesting messages in her Facebook account using images, video clips or songs (i.e., semantic digital artifacts demonstrating diversity in the modes used). Three digital artifacts were selected from her messages in her Facebook account.

From the study at Stage One (see Figure 1) of the three digital artifacts (see Table 2) and through observation of Angel's engagement on different web sites, the researcher noted that she had acquired the following abilities:

Table 2. The Study of Angel's Digital Artifacts

Digital artifact	Digital practices	Element 1	Element 2	Element 3	Element 4
Puppies as shark bait	<i>Mode:</i>	Writing	Writing	Image	Writing
	<i>Affordance:</i>	It described an idea (to persuade readers and urge appropriate action).	It described Angel's emotion.	It provides depictive information (She authenticated the claim with a newspaper cutting.)	It described an idea (to query).
	<i>Contextual link:</i>	<u>Elaboration:</u> The text (element 2) paraphrases the image (element 3).			
A psychological test	<i>Mode:</i>	Image	Writing	Writing	
	<i>Affordance:</i>	It provides epictive information (a serious look).	It describes an idea.	It describes an idea (the result of the test).	
	<i>Contextual link:</i>	Extension: The text (element 1) relayed the image (element 3).			
The handsome guys & the ugly guys	<i>Mode:</i>	Writing	Writing	Video	
	<i>Affordance:</i>	It describes Angel's emotion (funny).	It describes an idea (the comparison).	It provides information that demanded changes of time.	
	<i>Contextual link:</i>	<u>Explanation:</u> The text (element 2) made the video (element 3) more specific.			
	<i>Navigation</i>	An hyperlink			

- To engage with multi-mode elements, which recalled Jewitt and Unsworth's observations that the new literacies require users to engage with elements in different modes (Jewitt, 2006; Unsworth, 2008). Angle was competent in both recognizing and creating messages utilising elements of different modes, i.e., video, verbal narration, images and printed texts.
- To apply affordances of modes in meaning making. As Kress observes, different modes offer different potentials of meaning making (Kress, 2010).

### Images

According to Kress, one of the affordances of images is to provide depictive information. Angel applied a newspaper cutting to show the authentic aspect of the message (see Figure 2). In Figure 3, she used her "serious look" photo to reflect her friends' perception of her.

### Writing

Kress claims that writing affords the description of ideas. Angel urged readers of her message to tell more people (see Figure 2) about the news she had read that live puppies were used as shark bait; and she decided to change the image she was projecting of herself by using smiley faces to get to know more friends (see Figure 3).

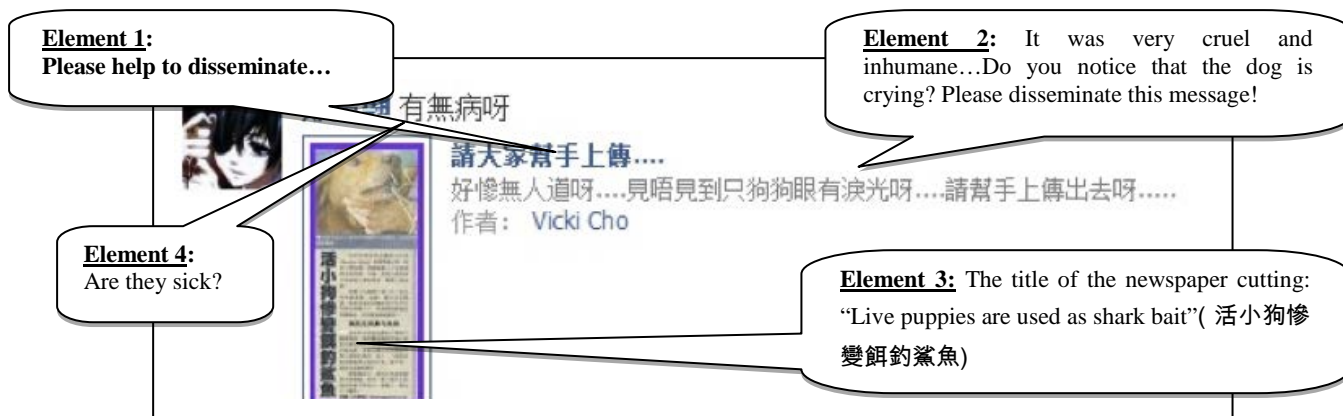


Figure 2. Angel's digital artifact with the title "Puppies As Fishing Bait."

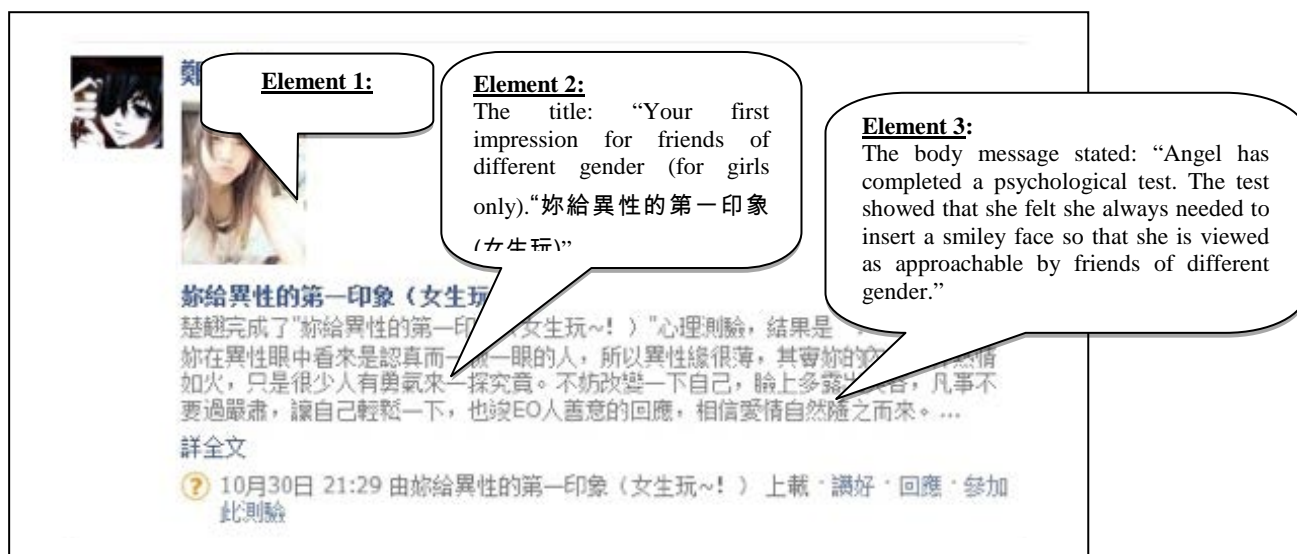


Figure 3. The digital artifact by Angel with the title, A Psychological Test

### Videos

Kress states that videos provide information that demand changes of time or space, or information that is perceived through sight and hearing (emotion, attitudes etc.). Angel applied a video clip that showed a series of images with different narration for each: "The Handsome Guys and The Ugly Guys" (<http://www.youtube.com/watch?v=vZhSvMRiXtE>). The narrations with illustrations compared how people perceived handsome and ugly boys performing a series of heterogeneous actions. For example: when a handsome boy reads a book, people would praise him as "a decent boy", but an ugly boy reading a book would be termed a "bookworm".

- To link elements of different modes meaningfully in layouts. One of the values of information on screens lies in its contextual links. The links integrate elements of information in different modes contextually in spatial or temporal layouts (Kress, 2010; Leeuwen, 2005); Angel demonstrated her understanding to elaborate on the title (writing) with the newspaper cutting (an image) as in Figure 2, and to complement the body message (writing) with a photo (an image) as in Figure 3.
- To navigate information on screens with hypertexts. Hypertexts differ from traditional texts. They afford readers interaction with text in a manner that is meaningful to them (Jonassen, 1986; Leeuwen, 2005; Martinec & Leeuwen, 2009). The researcher found that Angel could quickly adapt to the navigation potential and processing of digital texts of different modes.

In addition to the four digital practices that have been noted from the literature, another skill emerged from the Stage One study:

- The ability to assimilate digital functionalities in learning. Angel demonstrated her competency in searching for information on Facebook and, as mentioned, in using an online dictionary during the interview. This skill emerged independently. It did not derive from one the three of Angel's digital artifacts (Khoo & Churchill, 2012).

Digital texts appear in non-linear format on screens (Jewitt, 2006; Kress, 2010; Martinec & Leeuwen, 2009). The analysis of the digital artifacts in Table 2 showed that each digital artifact could be segregated into different elements. Angel engaged with these elements with the four digital practices mentioned in the literature: multi-mode, affordance of mode, navigation and contextual linking. The researcher observed some clear patterns emerging from Angel's engagement with these elements. There were two macro processes in her self-regulated engagement with these elements; she selected elements and integrated them with different links to form a semantic path of engagement.

Mayer's theoretical stance of cognitive multimedia theory guided the development of the initial framework of "Viewing and Representing Skills". The four digital practices as observed in the current research form the dimensions of the framework as in Table 3. The digital functionality skill was not observed in Angel's three digital artifacts. Therefore it was not listed in the initial framework of viewing and representing skills (see Table 3).

Table 3. The initial framework of Viewing and Representing Skills

The skills	Element Selection	Element Integration
	<i>Multi-mode</i>	<i>Navigation</i>
	<i>Affordance</i>	<i>Contextual Link</i>

Five mutually exclusive skills were identified. Furthermore, in Angel's three selected digital artifacts, the messages consisted of elements in different modes: i.e. images, sound, verbal narration, video and writing. The researcher categorized her as a 'Multi-mode representing practitioner.'

The second stage of the research was to determine how Angel extended her five viewing and representing skills into her English Language learning activities, which included formal schoolwork and independent schoolwork. The four skills were found extended from Stage One.

Two were observed in her formal schoolwork:

- Utilization of navigation skill (in her screen engagement she navigated with hyperlinks, multitasking and pop-up windows).
- Assimilating of digital functionality skill (she looked up words via Yahoo Dictionary, searched for information using the Yahoo Search function and "copied and pasted" when editing her writing).

Another two were found in her independent schoolwork:

- Creating messages with elements of multi-mode skill (in PowerPoint, she created messages with images and writing).
- Applying affordances of mode skill (she applied the affordances of modes in the same writing and images).

Note: The school did not promote the use of computer screens in assignments and independent schoolwork.

Interviewing of the teacher confirmed that the objectives of the lessons were to provide children with appropriate practices and knowledge as specified in the English Language curriculum. The teacher added that the lesson focused on tenses and comprehension practice. This suggested that Angel's formal English Language learning with digital text focused on the writing mode (i.e. meaning is transpired through mono-mode), while the content appeared in linear forms (the continuous engagement of readers in a fixed sequence). The interview with Angel's teacher showed that she was aware of two aspects of the viewing and representing skills:

- Assimilating digital functionalities  
Angel's teacher commented: "*Digital dictionaries can help children to learn independently...it is easy to manage children's work in terms of submission and marking...*" and
- Designing one's path of engagement  
The teacher stated: "*Learning through computers can allow children to learn actively.*"

### **Participant two: Winnie, a multimodal written-text representing practitioner**

Winnie is an eleven-year-old girl. She engaged with the Internet for only two days per week, each session being of two hours duration. The engagement comprised listening to songs, playing games and reading her blogs (she had a Yahoo blog account). She communicated with her friends through the blogs. Winnie preferred to use the writing mode in all her messages. She perceived that writing allowed her to express her feelings clearly to the reader.

In the first stage of research, four digital artifacts were selected from Winnie's out-of-school on-screen digital practices. Utilizing the data collection methods at Stage One as per Figure 1, the researcher observed that Winnie had acquired the same five viewing and representing skills. Although Winnie preferred to represent her messages in writing, the researcher noticed that she paid more explicit attention to the visual features of texts such as font size, formatting, paragraphing of main message and title. Winnie's written text provided meaning visually, more than linguistically. Messages in writing were treated as "resumes" (short statements). Winnie exhibited unique ways of presenting her messages in writing mode. The researcher classified her as a "multimodal written text representing practitioner."

In the second stage of the research, the researcher identified only 'screen navigation' and 'assimilating digital functionalities' as having been extended into Winnie's formal learning. She could use hyperlinks, multi-tasking and pop-up windows in her navigation, and regularly searched online for information in her attempts to complete her formal learning practices on screen. The "searching" digital functionality emerged to perform the process of "element integration" in the framework of viewing and representing skills.

The interviews with the teacher confirmed that the objectives of the lesson focused on conjunctions in English grammar and comprehension. During the interviews, the teacher also acknowledged the presence of two aspects of viewing and representing skills.

- Assimilating digital functionalities  
The teacher commented: "...the system can mark the children's schoolwork" and "...the online dictionary helps the child to look up the meaning of words."
- Designing one's path of engagement  
During the interview, the teacher acknowledged that: "the children are able to be more independent in learning."

### **Participant three: Nicole, a digital-savvy practitioner**

Nicole is a twelve-year-old girl. During the interview, the researcher observed her pulling out her smartphone to display her music and photos saved to the device. She gave the impression of routinely using her smartphone thus. At home, she accessed the computer twice per week. Each time she spent ninety minutes surfing the Internet. She used MS Office for some of her work, the school website to submit her schoolwork, MSN for communication and Yahoo to search for information related to her schoolwork. She also posted messages regularly on her blog at [www.Qooza.hk](http://www.Qooza.hk) and booked movie tickets online.

In the first stage of the research, five digital artifacts were selected from Nicole's out-of-school on-screen digital practices. The same five digital skills as observed in the previous cases were apparent in her engagement with the Internet and from her digital artifacts. However, she demonstrated different uses and appreciation of digital functionalities. The researcher termed her a "digital savvy practitioner."

In Stage Two - formal learning research - two skills were extended from Stage One:

- She navigated on the screen with hyperlinks.
- She was found to have assimilated digital functionality by editing her writing.

The digital functionality of "editing in writing" incorporated different types of macro process in the framework of the viewing and representing skills. "Editing in writing" was local and was an element selection.

When interviewed, Nicole's teacher identified the focus of the lesson as English Grammar, i.e., verb agreement and comprehension practice. Both were in writing mode. Although images and speech were present in the comprehension practice, both modes were found to be redundant factors in the students' attempt to answer the questions. However,

the teacher still perceived that images and speech could help children in comprehending the content. In the same interview, the teacher demonstrated her awareness of three aspects of the viewing and representing skills: The abilities of “designing one’s path of engagement,” “applying the multi-mode” and “screen navigation.”

#### Participant four: Wong, A Viewing Practitioner

Wong is eleven years old. She was a quiet and shy girl. She averaged 3 days per week using computers, mostly for finishing school homework and leisure. She used MS Word for her schoolwork. Each computer session comprised 90 minutes surfing the Internet and communicating on Facebook. She looked up words and searched for information online.

In the first stage of the research, five digital artifacts were selected from Wong’s out-of-school on-screen digital practices. Although her digital artifacts and engagement on-screen evidenced the same five viewing and representing skills as for previous cases, Wong was unique among the participants in that she did not create much informational content on-screen, preferring to view and read instead. The researcher classified her as a “Viewing Practitioner.”

The second stage of research indicated that she applied skills navigation and digital functionality in her formal learning activities. She navigated with hyperlinks and assimilated digital functionality by “copying and pasting” in her formal schoolwork and in her independent schoolwork, she applied skills multi-mode, contextual link and affordances of mode.

The interview with her teacher before and after the lesson identified the focus of the lesson as being on the writing mode. However, the teacher was aware of three aspects of viewing and representing skills: assimilating digital functionalities, designing one’s path of engagement and navigations.

### Results, discussion and an implication

In this study, five viewing and representing skills are identified (as in Table 4) in response to the first research question. Four of these skills - skills multi-mode, affordances of mode, contextual link and navigation - were derived from practices identified in the literature review before the data collection (Jewitt, 2006; Kress, 2010; Kress & Leeuwen, 2007; Leeuwen, 2005; Martinec & Leeuwen, 2009). An additional skill, digital functionality emerged from the study. These five skills were also found performing two different macro processes as shown in Table 4 below.

*Table 4. The framework of viewing and representing skills*

Macro Process	Element Selection	Element Integration
	<u>Multi-mode</u>	<u>Contextual Link</u>
	The skill to recognize and create elements of different modes to form information.	The skill to interpret and create contextual links (in spatial/temporal layouts) with different elements to form information.
<u>Skills</u>	<u>Affordances of mode</u>	<u>Navigation</u>
	The skill to apply and engage with the affordances of different modes in elements to form information.	The skill to move around a screen to integrate different elements to form information.
	<u>Digital functionality</u>	
	The skill to assimilate digital functionalities in elements to form information.	The skill to assimilate digital functionalities to integrate elements to form information.

A summary of how these skills are extended into formal school learning and independent schoolwork appears in Table 3. The second research question will be answered based on these findings.

In investigating how children extended their viewing and representing skills into formal schoolwork, the researcher found that “navigation” and “digital functionality” were utilized, while “multi-mode,” “affordances of mode” and

“contextual link” were not in evidence (see Table 5). The results suggest that the formal schoolwork focused on linguistic aspects of learning. Readers needed to peruse mono-mode text in a linear sequence to find meaning although the learning was performed using the screen medium. The studies from the four cases confirm that English language learning at school focuses on linguistic development (such as grammar and reading comprehension) as observed in classroom observations and the interviews with the teachers.

In the case of Nicole, although skill one (multi-mode) was extended as images and speech in her formal learning, both modes were found redundant in answering the questions. The result is consistent with the findings of other studies where images were found to be not utilized meaningfully in learning (Daly & Unsworth, 2011; Unsworth, 2008). By utilizing the effects of the distinctive affordances of images and speech (skill two) into the questions, the teacher’s intention to improve the level of children’s comprehension with multi-mode could be achieved.

It appears probable that the students’ use of the contextual links between images and writing mode (skill three) in their formal learning could be assisted if the new dimension of meaning making that is afforded on the medium of screens, i.e., the designs of learning resources- was introduced into the English Language curriculum (Kress & Bezemer, 2008).

Angel’s formal learning promotes mono-mode literacy (Kress, 2003; Unsworth, 2008) and linear text processing (Luke, 2003). Therefore, these would seem to suggest that skills two and three were not extended to her formal schoolwork.

*Table 5.* How primary school children extend viewing and representing skills learnt from their out-of-school digital practices into their formal and independent schoolwork.

Skill	Details	Out-of-School	Formal Schoolwork	Independent Schoolwork
Multi-mode	Writing	✓	✓	✓
	Images	✓		✓
	Video	✓		
Affordance of mode	Images	✓	✓	✓
	Writing	✓		✓
	Video	✓		
	Colour	✓		
Contextual link	Spatial	✓		✓
	Temporal	✓		
Navigation	Hyperlinks	✓	✓	
	Multitasking		✓	
	Pop-Up Windows		✓	
	Scroll Bar		✓	
Digital functionality	Multi-functionalities	✓	✓	

*Note.* “✓” denotes the skill is found to be evident in the area indicated in the title of the column.

However, the research found that the children intuitively extended “navigation” and “digital functionality” skills into their formal learning on screens (as in Table 5). Furthermore, feedback from the teachers indicated the fact that they were also aware of the skills. Mode and modal uses must be connected to the medium of distribution (Kress & Bezemer, 2008). Nevertheless, self-efficacy in digital functionality competency might not be sufficient to strengthen students’ knowledge and skills (Lai et al., 2012). If students are to acquire different learning digital functionality, the school should aim to provide them with appropriate knowledge and skills.

On the other hand, the study on the participants’ independent schoolwork revealed that neither Winnie’s nor Nicole’s school promoted digital text initiated school projects; Angel and Wong extended multi-mode, affordance of mode, and contextual link skills into their independent schoolwork. Besides creating their messages meaningfully with images and writing, they were also skilled at creating meaning with contextual links with non-linear spatial layouts. However, skills “navigation” and “digital functionality” were not in evidence in their independent schoolwork. This finding may be partially attributed to the criteria set by the teachers when the schoolwork was assigned. That is, there was no teacher-led promotion of schoolwork to be completed on screens with digital texts. Rather, the teachers requested that independent schoolwork be printed to hard copy.



A pedagogical implication arises from this research. Each participant had different preferences in relation to the ways in which they engaged with digital texts. At least four different viewing and representing practices were identified, and the respective participants were divided into categories as follows:

- Multimodal practitioner – one with viewing and representing skills who prefers to engage with elements of multi-modes formats
- Multimodal written text representing practitioner - one with viewing and representing skills that prefers to engage with writing mode format
- Digital savvy practitioner – one who is skillful in applying digital knowledge in day-to-day activities and practices
- Viewing practitioner - one who prefers to view information on screen, but not to represent their own ideas and messages thus

Four different learning practices that emerged in the current study are consistent with the different dimensions of learning styles as reported by (Silverman & Felder, 1988). This finding implies that education should adopt more individualized approaches to allow learners to design their own style and practice in learning. Other research also claims that different learning practices greatly affect screen-based learning performance (Huang et al., 2012).

## Recommendations

If digital literacies are to be effectively incorporated into primary school English Language learning, the following recommendations emerging from this study should be considered.

- The Primary School English Language Curriculum in Hong Kong should be revised to incorporate the viewing and representing skills, particularly the skills of “multi-mode”, “affordances of mode” and “contextual link”. The Government-mandated English Language Curriculum should not solely promote a linguistic view of literacy (with focus being on the mode of writing). This fails to connect the kinds of literacy required in the school with the “out-of-school worlds” of most students and other members of the greater community. The findings of this study have generated a deeper understanding that children habitually engaged in digital communication can be generally assumed to have acquired at least five viewing and representing skills. While traditional teaching approaches to the macro skills of listening, speaking, reading and writing (The Curriculum Development Council, 2004) are clearly still vitally important, the five skills identified in this report should also be incorporated into the Language classroom in order to embrace contemporary digital literacies. Further, teachers should be trained in the “multi-mode,” “affordance of modes” and “contextual link” skills.
- English language learning should promote media-mediated public pedagogy via popular culture modes such as blogs and social media. This pedagogy enhances learning on screens and promotes self-generated ideas (Richards & Jennifer, 2010). The research findings show that children extended different skills into both formal learning and independent schoolwork (see Table 5). Children did not extend “multi-mode,” “affordances of mode” and “contextual link” skills into their formal learning on screens. However, they extended these three skills into their independent schoolwork. It follows, then, that the focus on linguistic learning in a mono-mode model on screen might not optimize the benefit to learners. Children also did not extend “navigation” and “digital functionality” skills into their independent schoolwork, for which the teachers demanded hard copy printouts. Language learning should not be treated as a funnel into which all digital literacies must be squeezed through the narrow passage of current print-directed meaning representations. English language with digital literacies should expand the tremendous restrictiveness of the current practices. The media-mediated public pedagogy can bridge the shortcomings of both formal learning and independent schoolwork. Formal and independent schoolwork should enhance blogs, for example, where children can extend multi-mode, affordances of mode and contextual links into their formal schoolwork and navigations and digital functionalities into their independent schoolwork. It is likely that teachers would embrace the changes, since those interviewed acknowledged the skills that were afforded on screens.
- A self-regulated on screens learning approach should be promoted in English Language learning. The interviews with teachers demonstrated that they believed children learnt actively on screens. This research has pointed out that the two macro-processes of the viewing and representing skills need to be enhanced to incorporate different learning practices at school. Children do not necessarily learn in exactly the way the

teaching is set. The lack of specific direction for teaching about and with digital texts has been recognized (Walsh et al., 2007). New pedagogical strategies that support self-regulated learning, such as social media, computer games, and public pedagogy of technological spaces (Richards & Jennifer, 2010), which are drawing children away from traditional educational settings – must be promoted.

## Limitations

The research reported on in this paper had several limitations. Although the study focused on how children extend their out-of-school digital practices into English language learning, most of the digital artifacts deriving from their out-of-school practices were not in the English language. Therefore, the research findings may not be generalizable to English Language learning on screens. Further, different subjects might have different skills of focus (Huang et al., 2012). Future research could focus on the full range of school subjects and the essential digital skills relevant to each.

Doubt must be acknowledged as to the validity of inferences about the children's abilities to extend their skills developed from out-of-school digital practices to their independent schoolwork, since little such work was assigned. Further research could be conducted to focus on schools that do assign their students independent schoolwork on screen. Our study is focused on P5–P6 grade students (Primary Five and Primary Six); it should be highlighted that homework differs according to grades. Therefore, it may not be valid to make inferences about other grades on the basis of the current research results. Future study could focus on students of different age groups.

## Conclusions

The current multiple case studies research identifies at least six emerging viewing and representing skills that children have acquired from their out-of-school digital practices. The writing mode was found to be the focus of the participants' learning in their English Language classrooms. However, it was observed that the children have intuitively extended some of their viewing and representing skills into their formal learning and independent schoolwork.

Although several researchers have investigated aspects of digital multimodal texts in addressing new literacies for young children (Jewitt, 2008; Kress, 2003; Martinec & Leeuwen, 2009; Unsworth, 2001; Unsworth et al., 2005), the lack of specific direction in integrating digital texts into primary school English Language courses in Hong Kong is yet to be thoroughly explored. The current research provides a framework for the integration of digital texts into the primary school curriculum in Hong Kong, along with specific recommendations that emerged from the research findings and implications.

Future research could focus on the full range of subjects and the digital skills that might be essential in each subject, with the view to providing a set of specific guidelines for the respective subject teachers. While it is acknowledged that inferences about other subjects and grades should not necessarily be drawn on the basis of the current research, which is limited to primary school students of P5-P6 level, the results provide some groundwork and suggest implications for future research into language and literacy learning in primary schools.

## References

- Buckingham, D., & Willett, R. (2006). Is there a digital generation? In D. Buckingham (Ed.), *Digital Generations* (pp. 1-18). London, UK: Lawrence Erlbaum Associates.
- Connelly, J. (2008). Symbolic constructions in global public visuals: A pedagogic framework for critical visual literacy. In L. Unsworth (Ed.), *Multimodal Semiotics* (pp.159-170). London, UK: Continuum International Publishing Group.
- Cope, B., & Kalantzis, M. (Eds.). (2000). *Multiliteracies: Literacy, learning and the design of social futures* (pp.203-234). Melbourne, Australia: Macmillan.

- Daly, A., & Unsworth, L. (2011). Analysis and comprehension of multimodal texts. *Australian Journal of Language and Literacy*, 34(1), 61-80.
- Department for Children schools and families UK. (2009). The new primary curriculum. Retrieved from <http://webarchive.nationalarchives.gov.uk/20091207163458/>
- Gay, L. R., Mills, G. E., & Airasian, P. (2006). *Educational research - competencies for analysis and applications* (8th ed.). Columbia, OH: Prentice Hall.
- Gee, J. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment*, 1(1), 20. doi: 10.1145/950566.950595
- Huang, E. Y., Lin, S. W., & Huang, T. K. (2012). What type of learning style leads to online participation in the mixed-mode e-learning environment? A study of software usage instruction. *Computer & Education*, 58(1), 338-349.
- Jewitt, C. (2006). *Technology, literacy and learning - A multimodal approach*. Oxon, UK: Routledge.
- Jewitt, C. (2008). Multimodality and literacy in school classrooms. *Review of Research in Education*, 32(1), 241-267.
- Jonassen, D. H. (1986). Hypertext principles for text and courseware design. *Journal of Educational Psychologist*, 21(4), 269-292.
- Khoo, K. Y., & Churchill, D. (2012, March). *Case studies of children viewing and representing with digital text*. Paper presented at the 2012 Bangkok International Academic Conference, Bangkok.
- Kimber, K., & Claire, W. S. (2008). Assessing digital literacies: Can assessment ever be the same? In L. Unsworth (Ed.), *New Literacies and the English Curriculum* (pp. 328-352). London, UK: Continuum.
- Kress, G. (2003). *Literacy in the new media age*. London, UK: Routledge.
- Kress, G. (2010). *Multimodality- A social semiotic approach to contemporary communication*. Oxon, UK: Routledge.
- Kress, G., & Bezemer, J. (2008). Writing in multimodal texts: A social account of designs for learning. *Written Communication*, 25(2), 166-193.
- Kress, G., & Jewitt, C. (2003). *Multimodal literacy*. New York, NY: Peter Lang.
- Kress, G., & Leeuwen, T. V. (2006). *Reading images: The grammar of visual design* (2nd ed). London, UK: Routledge.
- Lai, C., Wang, Q., & Lei, J. (2012). What factors predict undergraduate students' use of technology for learning? A case from Hong Kong. *Computers & Education*, 59(2), 569-579.
- Leeuwen, T. V. (2005). *Introducing social semiotics*. Oxon, UK: Routledge.
- Luke, C. (2003). Pedagogy, connectivity, multimodality and interdisciplinarity. *Reading Research Quarterly*, 38(3), 356-385.
- Martinec, R., & Leeuwen, T. V. (2009). *The language of new media design-theory and practice*. Oxon, UK: Routledge.
- Mayer, R. (2005). Cognitive theory of multimedia learning. In R. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 31-48). New York, NY: Cambridge University Press.
- Merriam, S. B. (1988). *Case study research in education- A qualitative approach*. San Francisco, CA: Jossey-Bass Publishers.
- Merriam, S. B. (2009). *Qualitative research- A guide to design and implementation*. San Francisco, CA: Jossey-Bass.
- Ministry of Education Ontario. (2006). *The Ontario curriculum grade 1-8 (Language)*. Ontario, CA: Ministry of Education.
- MOE Singapore. (2010). *English language syllabus 2010 (Primary & Secondary)*. Retrieved from <http://www.moe.gov.sg/education/syllabuses/languages-and-literature/files/english-primary-secondary-express-normal-academic.pdf>
- National Curriculum Board Australia. (2008). *National English curriculum: Initial advice paper*. Retrieved from [http://www.acara.edu.au/k6.boardofstudies.nsw.edu.au/english/english\\_index.html](http://www.acara.edu.au/k6.boardofstudies.nsw.edu.au/english/english_index.html)
- Prensky, M. (2001a). Digital natives, digital immigrants. *On the Horizon*, 9(5), 1-6.
- Prensky, M. (2001b). Digital natives, digital immigrants: Do they really think differently? *On the Horizon*, 9(6), 1-9.
- Richards, L. F., & Jennifer, A. S. (2010). Shaping youth discourse about technology: Technology colonization, manifest Destiny, and the frontier myth in Facebook's public pedagogy. *Educational Studies*, 46, 503-523.

- Silverman, L. K., & Felder, R. M. (1988). Learning and teaching style. *Engineering Education*, 78(7), 674-681.
- The Curriculum Development Council. (2004). *English language curriculum guide (Primary 1-6)*. Retrieved from [http://www.edb.gov.hk/FileManager/EN/Content\\_2801/primary\\_1\\_6.pdf](http://www.edb.gov.hk/FileManager/EN/Content_2801/primary_1_6.pdf)
- Unsworth, L. (2001). *Teaching multiliteracies across the curriculum: Changing contexts of text and image in classroom practice*. Buckingham, England: Open University Press.
- Unsworth, L. (Ed.). (2008). *New literacies and the English curriculum*. London, UK: Continuum International Publishing Group.
- Unsworth, L., Thomas, A., Simpson, A., & Asha, J. (2005). *Children's literature and computer-based teaching*. London, UK: McGraw-Hill.
- Walsh, M., Asha, J., & Spranger, N. (2007). Reading digital texts. *Australian Journal of Language and Literacy*, 30(1), 40-53.
- Yin, R. K. (1994). *Case study research: Design and methods* (2nd ed.). Thousand Oaks, CA: Sage.

## Computer Enhanced English Language Tool for Students with Hearing Loss – A Bulgarian study

Milen Zamfirov<sup>1\*</sup> and Svetoslava Saeva<sup>2</sup>

<sup>1</sup>Department of Special Pedagogy and Speech Therapy, Faculty of Preschool and Primary School Education, St. Kliment Ohridski Sofia University, Sofia, Bulgaria // <sup>2</sup>Department of Pedagogy, Faculty of Pedagogy, Neofit Rilski South West University, Blagoevgrad, Bulgaria // milen\_zamfirov@abv.bg // s\_saeva@mail.bg

\*Corresponding author

(Submitted February 19, 2012; Revised June 28, 2012; Accepted October 22, 2012)

### ABSTRACT

This paper illustrates the results of teaching English to students with hearing loss using a computer program developed for learners defined as either hard of hearing or deaf. The research takes place at one of the three specialized schools in Bulgaria – “Prof. Dr. Decho Denev” Special Secondary School with Kindergarten for Children with Hearing Loss, located in Sofia City. The software was utilized for one academic year by students ranging from sixteen to twenty-three years of age. The program includes two video screens; one for lip-reading English phrases and one for interpreting them into Bulgarian Sign language with sentences classified into topics and pictures representing each phrase.

### Keywords

English teaching, Students with hearing loss, Computer literacy

### Introduction

There are over seventy million people with hearing loss throughout the world. They identify themselves within different communities, depending on whether sign language is their first language and on the type of communication environment - one consisting of hearing individuals or those with hearing loss. People with hearing loss who are willing to learn to communicate by speech, to become a part of the community of their hearing parents and relatives, are described as having a hearing loss. On the other hand, there are hearing impaired people that see themselves as representatives of another culture - holders of a different language - unlike those from the verbal language of the hearing population. These are identified with the group of the Deaf – with a capital “D” (Brown, 1995; Desselle, 1997). They are most often children of parents with hearing loss, who communicate mainly with other Deaf people and who do not want to be integrated in the hearing community (Campbell, 2001).

### Literature review

The communication process for people with hearing loss is characterized by the use of sign language, as well as with the inclusion of verbal communication. The majority of an unhearing population becomes proficient, to varying extent, in the official language of its native country (Qualls-Mitchell, 2002; Kent, 2003). The language of young hearing children develops in accordance to a certain system and structure, which is examined and discussed in Gillen’s study (2003). Other researchers draw a comparison between the norm for hearing children and for those with hearing loss and the processes involved in mastering the first language (McAnally et. al., 1987; Schein & Stewart, 1995).

The learning of a second or third language, by students with hearing loss, is a long and complicated process. The factors determining the proficiency level depend individually on the learner. It is also important to consider the significance of this information in the students’ everyday life. This presents the most important function of foreign language education – that of communication (Barnard & Glynn, 2003).

A number of research studies have been implemented, concluding that a different way of perceiving information exists among people with hearing loss, particularly among those in school environments (Andrews et. al., 2004). The teaching process has to be consistent with the specific needs of these students (Marschark, 1997). Part of one teaching method corresponding to these particular needs is Auditory Verbal Therapy (AVT). The treatment is an

important element of the students' integration into hearing society through its result – comprehensible speech. Its role is described in a longitudinal study by Dornan et al. (2010).

While learning their native tongue, sign or verbalizations, visualizing the information supports the social development of the children with hearing loss (Kent, 2003). The thinking, the ability to master, and the development of the child's personality are all related. (Marschark et. al., 1997, 2002).

Specific strategies and methods for the education of persons with hearing loss are used, which are not applicable in classrooms for students without such impairment (Stewart & Kluwin, 2001; Sanders, 1988). Alternative techniques and methods for education are necessary when children with hearing loss have to be educated in a class of hearing students (Watson et. al., 1999; Moores & Meadow-Orlans, 1990).

According to research findings, the opposite approach for educating deaf students, which uses mainly verbal speech, is a better one and offers superior strategies for the personal development of students with hearing loss (Hull, 1997; Mosely & Bally, 1996).

In Bulgaria, the foreign language teaching process is defined by law. The educational programs include compulsory abilities training in reading and writing, along with listening and speaking, yet the latter two skills are inaccessible and/or hardly realizable by students with hearing loss.

The syllabus in Bulgaria is similar to those of other countries. For example, in the United Kingdom (particularly in England and Wales), the language activities are divided into three areas:

- “Everyday activities”
- “Personal and Social life”
- “The World around us”

Moreover, there is a provision for development of the same four skills, which are included in the Bulgarian syllabus, namely:

- listening and comprehension
- speaking
- reading and comprehension
- writing

There are many more differences in the methods for teaching English language in the special schools for children with hearing loss as compared to mainstream schools. In Bulgaria, students with hearing loss are educated in both specialized, and conventional schools. When comparing these two, differences are found in:

- educational environment (hearing/with hearing loss classmates and teachers)
- domestic environment (students living with their families or far from them, at boarding schools)
- modality, used dominantly during the educational process (listening and recording/showing, explanation through visual models and examples)
- code of communication (verbal language/sign language)

## **Methodology**

The Methodology heading includes two sections:

- Description of the English teaching computer program for people with hearing loss and
- Analysis of the results from two special schools for children with hearing loss in Bulgaria (Sofia city and Plovdiv city)

### **Description of the multimedia CD-ROM “English language for people with hearing loss (beginners)”**

There is no effective Government-sanctioned policy for teaching English language to people with hearing loss, in Bulgaria. For that reason, a multimedia computer program; “English language for people with hearing loss

(beginners)” has been developed (see demo version at: <http://signlanguage-bg.com/bg/english.html>). It can be used for both group and individual work. It can also be utilized by hearing people who want to learn Bulgarian sign language.

The CD includes 220 phrases, divided into four subject areas, in accordance with the State educational requirements of the Bulgarian Ministry of Education, Youth and Science.

The phrases are presented in English, Bulgarian and Bulgarian sign language. The students gradually learn English on a beginner’s level, as they master and improve their computer skills. The user can navigate in the multimedia computer program via various commands and tools, designed for ease of use.

The CD is divided into three parts:

- Subject areas
- Various alphabets
- Authors

The following provides a more detailed explanation of each of these divisions:

### *I. Subject areas*

The subject areas are the main topics comprised of phrases for communication. Their number in each part varies. While choosing a certain phrase in the appropriate menu, two screens with video clips open up: the first one shows the sentence in Bulgarian sign language, where the gestures are accompanied by articulation and pronunciation in Bulgarian, while the second foreground, aiming at mastering the skill to lip-read the phrases, is presented in English.

The meaning of each phrase is revealed by a picture (not by a drawing, because the goal is to present the most natural view of the objects) that supports, by demonstration, the visual thinking and understanding of people with hearing loss.

All phrases are represented in three versions, with each being highlighted in one of the following colors:

- English (blue)
- Bulgarian transcription with stresses (red)
- Translation in Bulgarian (green)

The pronunciation is written in Cyrillic, rather than phonetic alphabet, as the correct pronunciation of some difficult consonants in the English language is beyond the abilities of people with hearing loss, sometimes eluding even the hearing beginner. The students with hearing loss encounter articulation difficulties even in their native Bulgarian tongue. The education is not focused on achieving correct pronunciation, but rather at the accomplishment of the main objective - making communication possible. The objective of passing the educational course presented on the CD is for the students to be able to communicate in written form, as well as to lip-read the familiar and learned phrases. For those who have well developed vocalization abilities in their native language, being able to speak effectively in English is also encouraged.

The phrases are divided into 4 subject areas:

#### **1. Person and communication**

Six themes are included:

- Introduction and presentation
- Family, relatives and friends
- Outlook and apparel
- Character
- Phone and Post
- Shopping

## 2. Everyday life

The section consists of three themes:

- School
- Home and daily regime
- Health and sports

## 3. The world around us

The following subjects are included:

- Native land and birthplace
- Nature (animals, plants and environment)
- Travelling and means of transport
- Countries and people
- Traditions, feasts and holidays

## 4. Activities

This section presents three subjects:

- Class activities
- Free time
- Expression of preferences and wishes

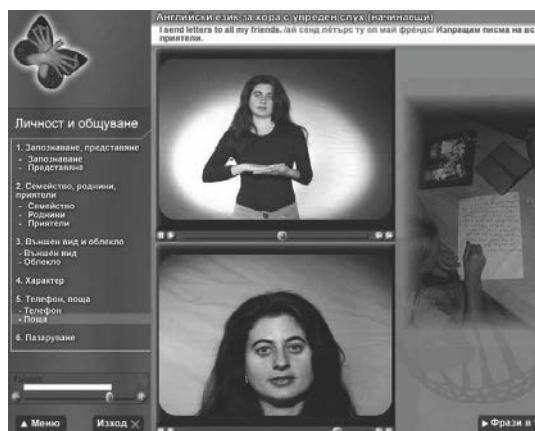


Figure 1. Screen from the unit: Person and communication

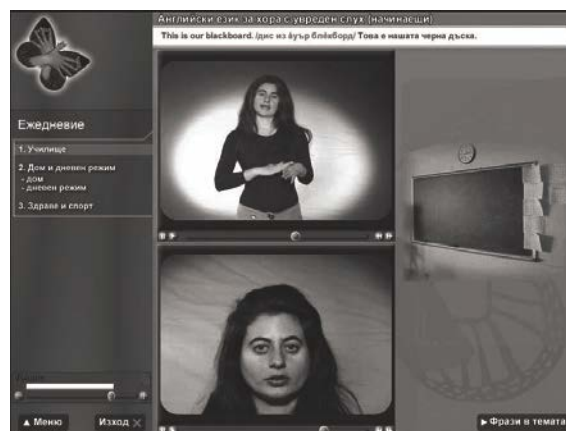


Figure 2. Screen from the unit: Everyday life



Figure 3. Screen from the unit: The world around us

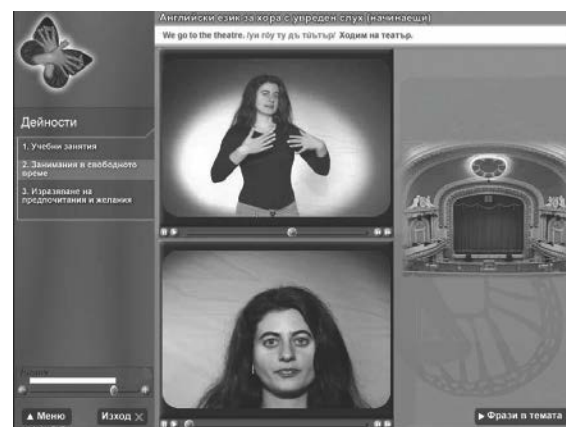


Figure 4. Screen from unit: Activities



## II Various alphabets

The section consists of one screen, showing a finger-spelled letter in sign language. All the alphabet letters can be seen – in English and in Bulgarian, and when selected, can appear larger, right from the video screen. Many Bulgarian students with hearing loss face difficulties distinguishing lowercase and uppercase signs in English and they confuse them with the Bulgarian ones, where the lowercase one is usually a copy of the corresponding uppercase letter.

There is also an examination of the English and Bulgarian alphabets, the purpose of which is to compare the graphemes and their correct pronunciation. Students with hearing loss find it difficult to master the pronunciation of English letters, either by themselves or included in linguistic construction. This is due to the significant differences between English and Bulgarian. For example, it is hard for people with hearing loss to realize that the letter “A” should be articulated “ei”, not “ah”, as it is in Bulgarian. Thus, they would pronounce the word *April* as /ahpril/, not as /eipril/ and by default, they would keep the stress on the second syllable as it is in Bulgarian, instead of the first. This persists amongst students from the special schools for children with hearing loss even after three years of learning English. Of course, skill development and progress depend upon the motivation of each individual student.

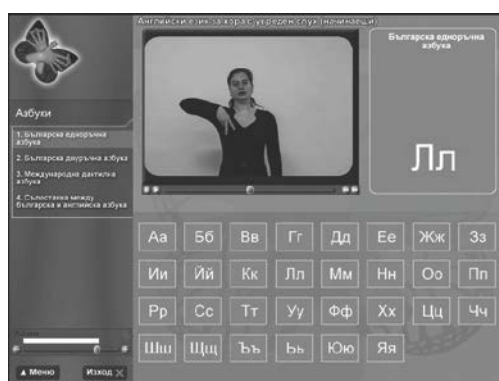


Figure 5. Screen from Various alphabets – the letter L in Bulgarian one-hand alphabet



Figure 6. Screen from Various alphabets – the letter L in Bulgarian two-handed alphabet



Figure 7. Screen from Various alphabets – the letter L in international manual alphabet

БЪЛГАРСКА АЗБУКА		АНГЛИЙСКА АЗБУКА	
ПИСАНЕ	ГОВОРЕНЕ	ПИСАНЕ	ГОВОРЕНЕ
Аа	/a/	* Aa	/ei/
Бб	/b/	* Bb	/bi/
Вв	/v/	* Cc	/ei/
Гг	/g/	* Dd	/di/
Дд	/d/	* Ee	/ai/
Ее	/e/	* Ff	/ef/
Жж	/zh/	* Gg	/dgi/
Зз	/z/	* Hh	/ei/
Ии	/i/	* Ii	/ai/
Йй	/i кратко/	* Jj	/dsei/
Кк	/k/	* Kk	/kei/
Лл	/l/	* Ll	/el/
Мм	/m/	* Mm	/em/
Нн	/n/	* Nn	/en/
Оо	/o/	* Oo	/oi/
Пп	/p/	* Pp	/pi/
Рр	/r/	* Qq	/koi/

Figure 8. Screen from Various alphabets – Comparison of graphemes of Bulgarian and English alphabets

## Analysis of the results from two special schools for children with hearing loss, in Sofia city and Plovdiv city

The English language instruction to students at “Prof. Dr D. Deney” Special Secondary School for Students with Hearing Loss, in Sofia, is conducted in accordance with the above described computer program. The experimental research groups consist of these students.

Teaching English to students from the Secondary Special School for Children with Hearing Loss “Prof. Dr. S. Belinov”, in Plovdiv, is performed using traditional methods. These students represent the control group.

All students from both groups are in 9<sup>th</sup>, 10<sup>th</sup> and 11<sup>th</sup> grades. There are 47 learners in the experimental group and 39 in the control group, accounting for a total of 86 participants with hearing loss. The makeup of both groups, regarding gender ratio and degree of hearing loss, is very similar. At the beginning of the study, the students’ Bulgarian language skills level (as their first spoken one) was assessed. The one major difference distinguishing the two groups is the presence of the computer English teaching program for learners with hearing loss in the experimental group and its absence in the other. All students in the experimental group were taught English personally by Head assistant professor Svetoslava Saeva, PhD and Head assistant professor Milen Zamfirov, PhD, in the course of a single academic year. The teachers used the English language for people with hearing loss (beginners) computer program, as well as traditional foreign language teaching methods. The control group was instructed by teachers who used only traditional foreign language teaching tools.

The study consists of three parts; words, sentences and text. In the Words section, there are ten words that a student must describe, defining what they look like and their purpose. The Sentences section is comprised of three sentences with increasing level of abstract meaning using the included words, ranging from concrete to abstract. For the Text section, students must choose two pictures out of eleven and develop a text, describing what they see in the picture. The students have to address all three sections in three languages. They have to write in Bulgarian and English and show signs using Bulgarian sign language, while the researcher marks the verbal equivalent of the signs in special files.

All indicators listed below are designed to measure the quantitative and qualitative values of the results, as well as to illustrate how learners recognize and interpret the meaning of the written words in English. The research showed that there are some students with hearing loss in both the experimental group (at the beginning of the academic year of the study, before the computer program was used in the teaching process) and the control group (at the beginning and at the end of the academic year), who believe that a Bulgarian word written with Roman letters is actually an English word.

The analysis was performed, in part, by gathering data individually from each participant. There is a file for each learner, which was created during the study process; in Bulgarian and English by the students themselves and in Bulgarian sign language by the researchers.

The analysis was conducted in accordance with the following criteria and indicators:

### **Analysis of first group (words)**

#### *Quantitative analysis*

- Criteria – Answers and no answers (presence/lack of reaction upon the word stimuli).  
Indicator – Number of answers (written words).  
Indicator – Number of no answer (written words).
- Criteria – Correctly and wrongfully written words.  
Indicator – Number of correctly written words and supportive indicator translation.  
Indicator – Number of wrongfully written words.

#### *Qualitative analysis*

- Criteria – Linguistic product resulting from the instruction to the task:  
Indicator – Correlation of the object to a class of other group.  
Indicator – Description of the object or its purpose.  
Indicator – Lack/others (e.g. lack of reaction upon the word-stimuli, the student rewrites the word, etc.).
- Criteria – Linguistic material used for description:  
Indicator – Single words and supportive indicator translation.  
Indicator – Separate sentences.  
Indicator – Connected text.  
Indicator – Lack/others (e.g. lack of reaction upon the word-stimuli, the student rewrites the word, etc.).

### **Analysis of second group (sentences)**

#### *Quantitative analysis*

- Criteria – Answers and no answer (presence/lack of reaction upon the word-stimuli).  
Indicator – Number of answers.  
Indicator – Number of no answer.
- Criteria – Correctly and wrong written words.  
Indicator – Number of correctly written words and supportive indicator translation.  
Indicator – Number of wrong written words.

#### *Qualitative analysis*

- Criteria – Linguistic material used for description:  
Indicator – Single words and supportive indicator translation.  
Indicator – Sentences.  
Indicator – Lack/others.

### **Analysis of third group (connected text)**

#### *Quantitative analysis*

- Criteria – Answers and no answer (presence/lack of reaction upon the picture stimuli).  
Indicator – Number of answers.  
Indicator – Number of no answer.
- Criteria – Correctly and wrong written words.  
Indicator – Number of correctly written words and supportive indicator translation.  
Indicator – Number of wrong written words.

#### *Qualitative analysis*

- Criteria - Linguistic material used for description:  
Indicator – Single words and supportive indicator translation.  
Indicator – Single sentences.  
Indicator – Connected text.  
Indicator – Lack/others.

The supportive indicator “translation” was used because of the agreement that words and sentences written in English will be considered correct if students offer other answers, including translation of the words, as well as a description of their transcription and writing of Bulgarian sentences in English letters. The last two indicators did not show statistically significant correlations and, as such, were not included in the data analysis.

Sixteen indicators describing the students and their environment are used, as follows:

1. gender
2. intelligence quotient distributed into two groups (below/above average level)
3. age
4. diagnosis
5. clinical reasons for deafness
6. participation in a speech-hearing rehabilitation course
7. number of years learning English
8. Geographic consideration (where the student lives)
9. number of years attending school (e.g., he/she is at the special school since first grade)
10. mobile phone usage (because of the text messaging using Roman and Cyrillic letters)
11. Internet use
12. auditory status of the parents (hearing/deaf)
13. level of parents' fluency in English and
14. highest grade completed

Two sets of students participated in the study:

*First set – students from the experimental group*

There are 47 participants in this cluster, all in either 9<sup>th</sup>, 10<sup>th</sup>, and 11<sup>th</sup> grade. They are located in the Secondary Special School with Kindergarten for Children with Hearing Loss, “Prof. Dr. D. Denev”, Sofia. The students can be divided into 3 subgroups based on their grades.

1<sup>st</sup> subgroup – students in 9<sup>th</sup> grade. The 1<sup>st</sup> subgroup consists of ten boys and six girls for a total of sixteen participants. All are between 16 and 22 years of age. Of this subgroup, two are hard of hearing, seven are with mild hearing loss, and seven are defined as deaf. Three of the participants have been diagnosed as having a genetic hearing impairment, while the others lost their hearing later in life. A total of five students had not passed a speech-hearing rehabilitation program during pre-school age. All of them studied English for the first year.

2<sup>nd</sup> subgroup – students in 10<sup>th</sup> grade. There are sixteen participants in the 2<sup>nd</sup> subgroup with ages ranging from 17 to 23. Four of them are hard of hearing, four are with mild hearing loss, and eight are considered deaf. Two of them have been diagnosed as having a genetic hearing impairment, with one of the students having attended speech-hearing rehabilitation classes. All of them are in their second year of studying English in school.

3<sup>rd</sup> subgroup – students in 11<sup>th</sup> grade. This subgroup includes 15 students, consisting of 12 boys and three girls. Their ages range from 18 to 22 years old. Five are hard of hearing, five are practically deaf and five are considered totally deaf. Two of these have been diagnosed with a genetic hearing impairment. Four of them have not passed a course for pre-school speech-hearing rehabilitation. All have been studying English language in school for three years.

*Second set – students from the control group*

There are a total of 39 participants in this cluster. It consists of persons with hearing disabilities from the 9<sup>th</sup>, 10<sup>th</sup>, and 11<sup>th</sup> grades of Special Secondary School for Deaf Children “Prof. Dr St. Belinov”, Plovdiv. Participants were selected in such a way as to match the distribution characteristics of the experimental group as closely as possible. One caveat has to do with students in the 10<sup>th</sup> grade, where the number of learners in the control group represented half of the number in the experimental group. The differences stem from the fact that in Plovdiv there is one class of hearing loss students, while in Sofia there are two.

The control group is divided into three sub-groups based on the participants’ current grade.

1<sup>st</sup> subgroup – students in 9<sup>th</sup> grade. There are 16 learners in the 1<sup>st</sup> subgroup, consisting of ten boys and six girls. Their ages range from 16 to 18. Four are defined as hard of hearing, with 12 defined as deaf. A total of six report a genetic reason for their deafness, while the other 10 report losing their hearing sometime after birth. All have passed a course for preschool speech and hearing rehabilitation and all are in their first year of studying English language at school.

2<sup>nd</sup> subgroup – students in 10<sup>th</sup> grade. There are eight participants in the 2<sup>nd</sup> subgroup, consisting of four girls and four boys. Their ages range from 17 to 20 years. Two of the students have been diagnosed as genetically deaf. One has not passed a course for preschool speech-hearing rehabilitation and all are in their second year of studying English language.

3<sup>rd</sup> subgroup – students in 11<sup>th</sup> grade. This group consists of 15 participants, 12 of which are boys and three are girls. Their ages range from 18 to 26 years. Six of the participants are hard of hearing and nine are considered deaf. Two of those considered hard of hearing have been diagnosed with a genetic cause. All have passed preschool speech and hearing rehabilitation and all have studied English for the three years in school.

Total of 86 students, from two special schools, participated in the research. Their information is distributed by three indicators in Table 1.

*Table 1. Data about the participants in the research*

<b>Indicator</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Standard deviation</b>
Age	16	23	18.67	1.61
Intelligence quotient	58	120	88.49	13.09
The student attends school – number of years	1	18	10.19	2.984

The age of the students varies from 16 to 23, with a mean of 18.67. The intelligence quotient varies from 58 to a 120, where the average is 88. The students have attended school from one to 18 years; mean – 10 years. The distribution by gender is presented in table 2.

*Table 2. Distribution of the participants by gender*

<b>Indicator</b>	<b>Number</b>	<b>Percent</b>
Boys	55	64
Girls	31	36
Total	86	100

There are more male than female participants.

The interactions between the sixteen indicators (gender, Intelligence Quotient, etc.) was studied with the statistics method of Multidimensional Hierarchical Log-linear analysis. It is the equivalent to the Chi-square procedure for two-dimensional cross-tables. It is often preferred, if the total number of the examined persons is small, or if the cells count with the expected (theoretical) frequencies smaller than 5 exceeds 10 %.

K-way tests of significance – this is the summary of the k-way interaction tests, where k= number of factors, were checked. The SPSS procedure of Hierarchical Log-linear Analysis works with  $k \leq 10$  factors (indicators) at a time.

Partial Associations also were checked – they give the lower order (e.g., 9, 8, ..., 2-way) tests of independence.

Significant Main effect (interactions) and Partial association between more than two indicators were not found.

- Significant correlation between the gender of the students and the grade they are currently in was not found.
- Significant correlation between gender and IQ was not found.

There is one significant correlation between the diagnosis and belonging to the control or experimental group. Only one person from the control group has mild hearing loss, while in the experimental group there are 16 persons with such diagnosis.

### **Integral indicators**

The correctness of using integral indicators, by summarizing the results of the fifteen indicators utilized in the three series, was checked with the reliability coefficient - Cronbach's "Alpha." The Alpha coefficients are very high for the sectors Bulgarian and English, and medium high for the Sign language.

There are no errors registered for the 'incorrectly written' words indicator. The reason for this absence is also the lack of correct answers, because the set of hard of hearing persons, who do not use sign language in communication with hearing interlocutor, prefer to talk. There is also a group of students with mild hearing loss and deafness who prefer to dictate their opinion, using the verbal speech, because the information they provide is directed to a hearing interlocutor. Additionally, incorrect answers cannot be found in the Sign language sector, because all the examined persons are fluent in it.

From all the results mentioned above, the following conclusion can be stated:

The possibility for usage of summarized, integral indicators in all of the three series is formally confirmed:

- In the sign language sector ‘official zeros’ appear because some of the students with different level of hearing loss refuse to use gestures in communication with hearing interlocutor, as well as because those of them who use it for the communicative purposes do not make ‘wrong’ gestures.

## Quantitative analysis

### Descriptive statistics of the integral indicators: English language sector

In series one, four statistically important differences by indicators from the English language sector can be noted.

In series two, four significant differences by indicators pointing the linguistic competence can be found (one of them is the supportive indicator “translation”). The most important difference between the control and the experimental groups is found in the indicator number of reactions in the English language sentences. The experimental group shows on average 3 reactions, which is approximately seventy times higher than the one of the control group, which is estimated as 0.09.

In series three, there are four statistically important differences in the mean result between the control and the experimental group. They refer to the linguistic competence.

The statistical significance of the differences was registered from the T-test for comparison of the mean values in the random rows.

The results show the following: In the experimental group, as opposed to the control group, can be observed one statistical tendency and seven qualitative indicators in the English language sector.

## Quantitative analysis

Additional calculations of two supportive indicators are made:

- Number of results above the average level of the group in English language sector (linguistic competence English language sector)
- Number of results above the average level for the group in all the sectors – common linguistic competence.

The necessity of these indicators comes from the decision to measure the influence of the demographic characteristics of the children and those of their environment, as well as from the attempt to check the hypothesis to what extent the fluent knowledge of one or two languages helps in mastering a second (third) one.

For this purpose, the results from all indicators are standardized through the standard Z-marks of the procedure descriptive statistics (where average equals zero and standard error equals one). Thus, received standard values are transformed in two power values, in accordance with the criteria selected below and above the average.

In figure 9, the y-axis presents the number of participants who have given correct answers above the average level and the x-axis shows the indicators (described precisely in the text above).

In the experimental group, the topics (subgroups) are two: for 11 and 15 indicators above the mean. The control group does not contain frequencies above 16 indicators above the average for the group, while in the experimental one, 20 indicators above the average are with frequency of four percent.

Figure 10 shows the percentage allocation for the indicator for linguistic competence – English language sector.

In figure 10, the y-axis presents the percentage distribution in both groups above the mean level and the x-axis show the linguistics quantitative and qualitative indicators (described precisely in the text above).

The results from the experimental group are comparable. The sum of one additional indicator forms two topics (two sub-groups of students). One of them is for three indicators above the average with the frequency of 49 percent, while the second is for six indicators above the average with the frequency of 17 percent.

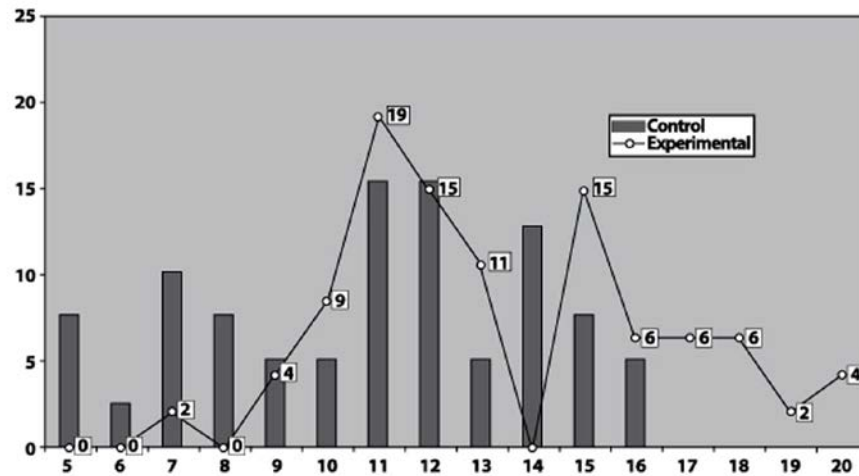


Figure 9. Frequency distribution of the number of results above the mean in both groups

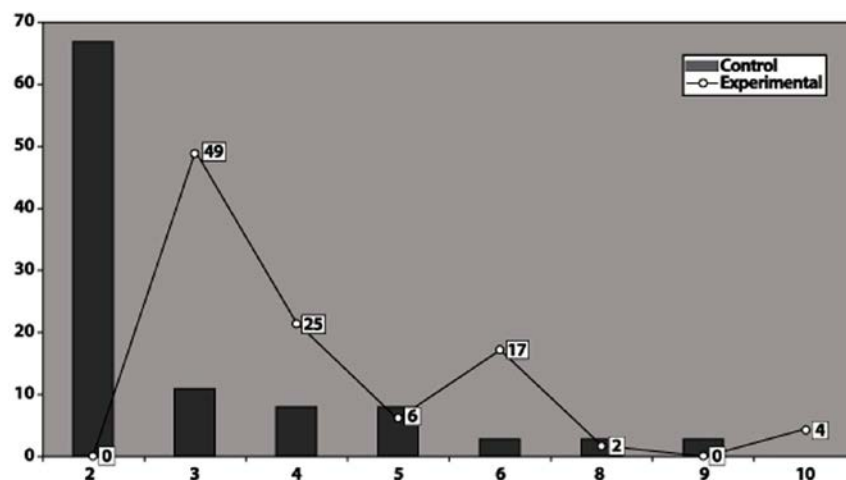


Figure 10. Percentage distribution of the indicator for linguistic competence – English language sector

These data point to better results demonstrated by the experimental groups. Verification of the process can be made with the formal statistical procedure T-test.

### Descriptive statistics for the integral indicators in the English language sector

In series one, one statistically significant difference and one tendency by indicators from the English language sector can be observed. One of them is an indicator for linguistic competence, and the other is for the absence of linguistic competence. The data is represented by figure 11.

The frequency distribution on figure 11 shows that the allocation of the incidences in the experimental group has two peaks (two sub-groups with different results) – the first one has zero sum of points for linguistic competence, while the second has four points.

Three peaks (two sub-groups with different results) are formed in the control group; the first has zero sum of points, the second has five points, and the third – eight points.

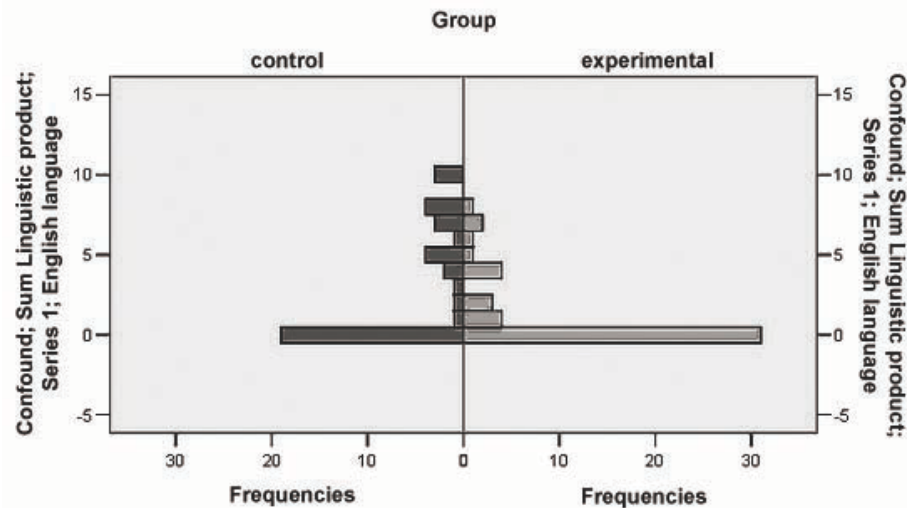


Figure 11. Frequency distribution for presence and absence of linguistic competence the examined persons have in English language

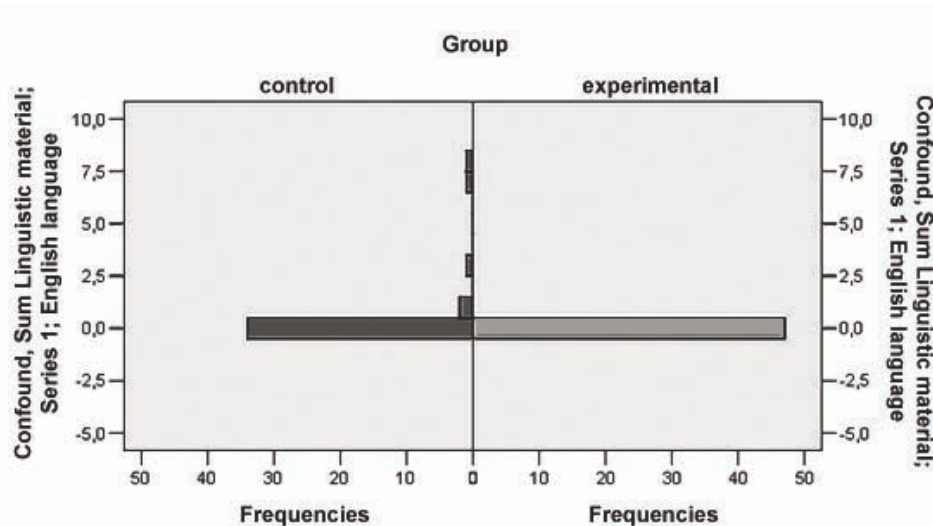


Figure 12. Series 1, criteria 2 – linguistic material – confounds

The statistical significance of the differences was formally registered with the T-test for comparison of the mean values of the random rows. In series one, four statistically important differences can be noticed by the following indicators:

- two from the English language sector
- two from the Sign language sector

Three of them are indicators for linguistic competence and one for the absence of such.

The experimental group shows higher results, as compared to the control group, by ten indicators – five statistically significant differences and five tendencies.

It can be summarized that, in the qualitative analysis, five statistically significant differences and five tendencies between the experimental and control groups were discovered.

The following conclusions can be drawn:



- The qualitative analysis also confirms the statistical significance of the highest results in the three series for the experimental group, compared to the control one.
- According to the number of indicators above the average for the group – the common linguistic competence – the experimental group shows 2,58 results more.
- According to the number of indicators above the average for the group – linguistic competence – the experimental group- English language sector shows a mean of one, three results more.

## **Impact of bi- and trilingual status upon the results**

Verifying the influence, which the proficiency in one or more languages has upon mastering another one, is achieved through the Factor analysis method. Markers are used (special indicators with certain qualities, which facilitate the interpretation, depending on whether they are joined to a concrete factor or not). All the factors are interpreted as relatively independent sub-groups consisting of conjoined indicators.

Only in series one, the results from the sign language sector have sufficient diversity in answers, in order to be included in the analysis. For this reason, only the results from series one are analyzed.

The summarized indicators are used as markers:

- Common linguistic competence
- English language competence

The primary results showed only a conditional validity of the factor model (Kaiser-Meyer-Olkin Measure of Sampling Adequacy = 0,50). That is why the individual adequacy of each indicator of the model was checked.

What is seen is that the model can be improved through distraction of the variables with individual adequacy under 0,50. After the consecutive distraction of these indicators from the sample, a total of six indicators remained (excluding the markers).

These indicators form a sample with satisfying reliability (Kaiser-Meyer-Olkin Measure of Sampling Adequacy = 0,598). A three-factor sample (three sub-groups of conjoined indicators) explaining 71 % of the variation - good sample for practical purposes was received.

In the first factor, the influencing coefficients (importance of the indicator in the factor) vary from –0,82 to plus 0,80. This factor can be called “success in the English language sector – cautiousness”, for the English language indicators have the highest absolute values of 0,84 and 0,43. It explains 30 % of the variation.

Because the lack of mistakes indicates certain success, the word “cautiousness” found its place in the name of the factor.

- The interrelation between the Sign and English language sectors demonstrates that the absence of mistakes in one of the languages is accompanied by the absence of mistakes in the other one.

The second factor explaining 27 % of the variation is related with success in the Sign language sector.

The success in the sign language sector is achieved through:

- High results by the indicator in the sum of the used linguistic material description/material and
- Low results by the indicators confound.
- Both for linguistic product criteria and for
- The linguistic material used.

At the same time, poor interrelation with the Bulgarian language sector can be noticed – a coefficient of minus 0,20.

- It comes to prove once again the founding in the English language sector – the success achieved in one of the languages is accompanied with a decrease of mistakes in the other.

The third factor explains thirteen, seven of the whole variation and concerns the success achieved in the Bulgarian language sector.

Summary:

- Two interrelations between sign and English languages are found, which leads to the premise that the success of one of them is accompanied with less mistakes in the other.
- The success in the Bulgarian language sector implies less mistakes in Sign language (two criteria) and in English language (one criteria).
- Sign language has an impact on learning English.
- English language demonstrates influence on becoming proficient in Sign language.
- Bulgarian language has an effect on Sign and English languages.
- The interrelations are not symmetric – the extent of impact Sign language has upon mastering English is -0,20 (absence of mistakes). The opposite dependency – the influence of English language over Sign language is -0,25.
- English and Sign language sectors do not affect the Bulgarian one.

## Conclusions

Teaching English as a second language to students with hearing loss is a challenge. However, an increasing number of teachers, both from specialized and mainstream schools, have to participate in this process. There are some false perceptions in Bulgarian high school students with hearing loss about the essence of English language (e.g. writing Bulgarian words in Roman alphabet).

We can draw four conclusions based on the study described in the article:

- The application of computer based teaching strategies with students with hearing loss in English classes is highly effective. There are many visual stimuli and manual commands which increase the user-friendliness of these strategies for students with hearing loss, during the teaching-learning process.
- Teaching English via “English language for people with hearing loss (beginners)” contributed to enhanced and accelerated mastery of the English syllabus and the long-term retention of the obtained knowledge.
- The study results showed a tendency for acquiring English language skills better with the computer program than with traditional foreign language teaching methods.
- Application of variety of methods and approaches is required in the educational process for achieving long-term results.

Language teaching is of great importance to children with hearing loss. It gives them pathways to communication with the hearing world.

The results of the research demonstrate the participants’ success in learning English language, as their second or third language, via a computer program. Instructors of students with hearing loss should be encouraged in using computer programs in the teaching process, not only when English language is concerned, but all other school subjects as well.

## Further research

Interactive computer programs are highly effective in teaching students. This effect doubles when educating learners with hearing loss.

There are many perspectives on the topic of teaching students with hearing loss another language. This paper presents the language focus on that issue, but it can also be part of psychological study, as well as of an AVT research.

## References

- Andrews, J., Leigh, I., & Weiner, M. (2004). *Deaf people. evolving perspectives drom psychology, education and sociology*. London, UK: Pearson Education Inc.
- Brown, P., & Gustafson, M. (1995). *Showing sensitivity to deaf culture*. Rochester, NY: National Technical Institute for the Deaf.

- Campbell, P. (2001). A Struggle for identity. *Humanities*, 22(2), 22-28.
- Desselle, D., & Pearlmutter, L. (1997). Navigating two cultures: deaf children, self-esteem, and parents' communication patterns. *Social Work in Education*, 19(1), 23-30.
- Dornan, D., Hickson, L., Murdoch, B., Houston, T., & Constantinescu, G. (2010). Is auditory-verbal therapy effective for children with hearing loss? *The Volta Review*, 110(3), 361-387.
- Gillen, J. (2003). *The language of children*. London, UK: Routledge.
- Barnard, R., & Glynn, T. (Eds.) (2003). *Bilingual children's language and literacy development*. Clevedon, England: Multilingual Matters.
- Hull, R. (1997). *Aural rehabilitation. Serving children and adults*. London, UK: Singular Publishing Group, Inc.
- Kent, D. (2003) *American sign language*. St. Louis, MO: Turtleback Books.
- Marschark, M. (1997) *Psychological development of deaf children*. Oxford: Oxford University Press.
- Marschark, M., Siple, P., Lillo-Martin, D., Campbell, R., & Everhart, V. (1997). *Relations of language and thought. The view from sign language and deaf children*, Oxford: Oxford University Press.
- Marschark, M., Lang, H., & Albertini, J. (2). *Educating deaf students: From research to practice*. Oxford, UK: Oxford University Press.
- McAnally, P., Rose, S., & Quigley, S. (1987). *Language Learning Practices with Deaf Children*. Austin, TX: Pro ed.
- Moores, D., & Meadow-Orlans, K. (1990). *Educational and developmental aspects of deafness*. Washington, DC: Gallaudet University Press.
- Mosely, M., & Bally, S. (1996). *Communication therapy: An integrated approach to aural rehabilitation*. Washington, DC: Gallaudet University Press.
- Qualls-Mitchell, P. (2002). Reading enhancement for deaf and hard-of-hearing children through multicultural empowerment. *Reading Teacher*, 56(1), 76-84
- Sanders, D. (1988). *Teaching deaf children. Techniques and methods*. Boston, MA: Pro ed.
- Schein, J., & Stewart, D. (1995). *Language in motion. Exploring the nature of sign*. Washington, DC: Gallaudet University Press.
- Stewart, D., & Kluwin, T. (2001). *Teaching deaf and hard of hearing students. Content, strategies, and curriculum*. Boston, MA: Pearson.
- Watson, L., Powers, G., & Powers, S. (1999). *deaf and hearing impaired pupils in mainstream school*. London, UK: David Fulton Publishers.

## Effects of Touch Technology-based Concept Mapping on Students' Learning Attitudes and Perceptions

Gwo-Jen Hwang<sup>1\*</sup>, Chih-Hsiang Wu<sup>2</sup> and Fan-Ray Kuo<sup>3</sup>

<sup>1</sup>Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taiwan // <sup>2</sup>Department of Information and Learning Technology, National University of Tainan, Taiwan //

<sup>3</sup>Department of Information Management, National Sun Yat-sen University, Kaohsiung, Taiwan // gjhwang.academic@gmail.com // paulwu.mail@gmail.com // revonkuo@gmail.com

\*Corresponding author

(Submitted February 15, 2012; Revised August 21, 2012; Accepted August 21, 2012)

### ABSTRACT

Concept maps have become a widely used educational tool around the globe. The advancements in computerized interface technologies have enabled even more alternatives for using concept maps in teaching and learning. This study investigates the effects of two different touch technology-based concept mapping interaction modes on students' learning achievements and learning attitudes in a natural science course, as well as their degree of acceptance of using concept maps to learn. Ninety two sixth graders were randomly divided into three groups. Experimental Group One was taught using the Interactive Whiteboard (IWB)-based concept mapping approach, Experimental Group Two learned with the touchscreen-based concept mapping approach, while the control group learned with the traditional paper-and-pencil-based concept mapping approach. The experimental results show that, in terms of learning attitudes toward the natural science course and the degree of acceptance of using concept maps to learn, the students were significantly more positive about the two touch technology-based interaction modes than they were about the traditional paper-and-pencil mode.

### Keywords

Concept map, Interactive whiteboard, Touch technology, Learning attitude, Technology acceptance model

### Introduction

This study investigates the effects of using two different touch technology-based interaction modes of computerized concept mapping on students' learning performance. A concept map is a diagram showing the relationships among concepts. It is a graphical tool proposed by Novak and Gowin (1984) for organizing and representing knowledge (Ruiz-Primo, Shavelson & Schultz, 1997). It has also been recognized as a meta-cognitive tool which can empower learners to monitor and control their cognitive progress (Novak, Gowin & Johansen, 1983; Novak, 1990).

With the rapid development of information technology, concept mapping is no longer confined to the production of paper-based drawings. Instead, students are able to modify their own concept maps conveniently via computerized concept mapping tools. The advancement of touch technologies has further enhanced the process by offering easier-to-use interfaces that facilitate the use of concept maps in teaching and learning. For example, Interactive Whiteboards (IWBs) allow instructors to show a variety of content to students which can promote their interest in inquiring into proposed questions (Hwang, 2003; Hung, Lin & Hwang, 2010). Research by Smith (2000) indicated that IWBs are appropriate for any curriculum and for all ages, and can present visualized materials such as text annotation, hidden objects, fast-moving objects and image scaling, etc. It has been found that use of these features can increase teachers' autonomy and flexibility (Kennewell, 2005; Xu & Moloney, 2011).

Based on the aforementioned motivation, this study investigates the effects of different touch technology-based concept mapping approaches on both students' learning attitudes toward a natural science course and their degree of acceptance of the learning approaches. Two experimental groups and one control group were assigned to the three different strategies, that is, the IWB-based concept mapping approach, the touchscreen-based concept mapping approach, and the paper-and-pencil-based concept mapping approach. The proposed research questions are as follows:

- Do the students who learn with the touch technology-based concept mapping approaches show significantly better attitudes toward natural science in comparison with those who learn with the traditional approach?
- Do the students who learn with the touch technology-based approaches show significantly better acceptance of use of the concept maps for learning in comparison with those who learn with the traditional approach?

## Literature review

### Computerized concept maps

Concept mapping for meaningful learning as proposed by Novak and Gowan (1984) is a theory of instruction based on the meaningful learning principles proposed by Ausubel (1963). Novak and Gowan (1984) aimed to utilize "concept maps" to represent meaningful relationships between concepts and propositions. A concept map can be regarded as a "visual road map" showing the pathways connecting the meanings of concepts (Novak, 1996).

In the past decades, concept maps have been used to help learners reflect on, organize and adjust their given or learned concepts in specific domains (Roth & Roychoudhury, 1994). They have also served as a form of measurement for assessing students' learning achievements (Markham, Mintzes, & Jones, 1994; Schmid & Telaro, 1990). Some previous studies have shown that concept map-based approaches can be used to diagnose learners' knowledge structures and misconceptions (Hwang, 2003; Novak, 1990; Panjaburee, Hwang, Triampo & Shih, 2010). McClure, Sonak and Suen (1999) further summarized concept maps as having four potential applications in science education: (a) as a learning strategy, (b) as an instructional strategy, (c) as a strategy for planning curricula, and (d) as a means of assessing students' understanding of science concepts.

The traditional pen-and-paper-oriented approach to concept mapping has several drawbacks, the foremost among which is that concept maps produced in this way are difficult for students to modify and for teachers to evaluate, which often reduces their effectiveness. Moreover, with the paper-and-pencil approach, the development process of the concept maps cannot be recorded, and hence the teachers can only evaluate students' learning performance based on their final outcomes. Therefore, researchers have developed computerized concept mapping tools to cope with these problems; one example of such a tool is CmapTools, developed by the Institute for Human and Machine Cognition (IHMC) of the Florida University System (Novak & Cañas, 2006). In addition, researchers have created more facilities to aid students in developing quality concept maps; for example, Wu, Hwang, Milrad, Ke and Huang (2012) developed a mechanism for evaluating students' concept maps and giving them instant feedback to help them make improvements.

Owing to the popularity of computer networks, some researchers have developed computerized concept mapping systems in web-based environments, which allow learners to collaboratively develop concept maps via network communications. The study of Khamesan and Hammond (2004) showed that students who learned with web-based concept mapping systems learned as effectively as those who learned with the face-to-face collaborative concept mapping approach. Kwon and Cifuentes (2009) further reported that computerized collaborative and individual concept mapping approaches had equally positive effects on students' learning achievements; however, students who learned collaboratively created significantly higher quality concept maps than those who learned individually, indicating that deeper conceptual understanding could be gained via collaborative learning.

In recent years, the advancement of mobile technologies has further engaged researchers in developing concept map-oriented mobile learning systems (Hung, Hwang, Su, & Lin, 2012). For example, Hwang, Wu and Ke (2011) proposed an interactive concept map-oriented approach for supporting mobile learning activities in the field for an elementary school natural science course. Their experimental results showed that the proposed approach not only enhanced learning attitudes, but also improved the learning achievements of the students. From the literature, it can be found that concept mapping approaches have not only been widely adopted by educators and researchers, but have also been implemented on different learning platforms, implying the potential of such approaches.

Currently, computerized concept mapping tools such as MindManager, Inspiration or FreeMind can be used to create concept maps with a keyboard and mouse. However, researchers have indicated that this approach suffers from two problems: first, it usually requires users to follow particular task structures (e.g., two concept nodes must be created before a proposition can be made), which could distract users from their primary tasks; second, it becomes inefficient and ineffective to use a keyboard and mouse on specific devices that support more direct interaction, such as touch-based computers, because such devices only offer a virtual keyboard and gesture-based interaction that differs from mouse behaviors (Morgan & Butler, 2009; Rick & Rogers, 2008; Ruiz-Primo, Shavelson & Schultz, 1997).

Accordingly, touch technologies provide flexible ways for learners to interact with instructional contents, and thus support free and enjoyable interactions that contrast with the well-established practices of paper-and-pen-based

interaction (Hollan & Hutchins, 2010). That is, those new technologies have the potential of benefiting learners in allowing them to discuss, annotate, and manipulate shared digital artifacts in a more direct and enjoyable way.

### **Touch technologies for computers**

The term “touch technology” refers to the development of digitalized interfaces that are able to detect the presence and location of a touch within a display area; examples of such interfaces are touchpads, touchscreens, and Interactive WhiteBoards (IWBs). Touchscreens enable people to interact directly with what is displayed rather than indirectly with a mouse or touchpad. They have played a prominent role in the design of digital appliances such as personal digital assistants (PDAs), satellite navigation devices, mobile phones, and video games (Hwang, Tsai, Chu, Kinshuk, & Chen, 2012; Wu, Hwang, Su, & Huang, 2012).

Researchers have indicated that touch technologies allow users to operate computer systems more easily than ever before, making them highly attractive to teachers and students (Romeo, Edwards, McNamara, Walker, & Ziguras, 2003; Yu, Zhang, Ren, Zhao, & Zhu, 2010). For example, the British government not only invested a considerable sum in installing Interactive WhiteBoards (IWBs) in elementary and junior high schools from 2002 to 2004, but also initiated the "National Whiteboard Network" project. In a similar move, the Hong Kong Education and Manpower Bureau started promoting the use of IWBs in elementary and junior high schools in 2004.

The study of the implementation and impact of these projects has found that in certain subjects, the more experience the teacher has of using the interactive whiteboard, the greater the likelihood of positive attainment gains in mathematics and natural science courses for the students (Somekh et al., 2007). Goodison (2002) investigated primary schoolchildren's awareness of the linkage between ICT and the way they learn within the context of a school that has been particularly successful in integrating IWB into its curriculum. He also stressed the need to meet the mission requirements with appropriate software. The studies of Wood and Ashfield (2008), and Dhindsa and Emran (2006) both found that in an IWB learning environment combined with collaborative learning activities, teachers should offer students more opportunities for communication, interaction and cooperation to construct and rework their knowledge structures.

To sum up, touch technology has potential in educational applications; therefore, it is worth studying the effects of different forms of touch technology used as educational tools on the learning performance of students.

### **Research design**

The purpose of the current study is to investigate the effects of different touch technology-based concept mapping strategies on the learning attitudes and learning achievements of elementary school students in a natural science course. To obtain accurate research results, three variables which could affect the experiment needed to be controlled, namely the instructor, learning time and learning content, as shown in Figure 1.

### **Participants**

The study was conducted in an elementary school in southern Taiwan. A total of 92 sixth-grade students (12-13 year olds) from three classes participated in the experiment. Two of the classes were assigned to be the experimental groups, and the other was the control group. Experimental Group One, consisting of 31 students, learned with the IWB-based concept mapping approach. Experimental Group Two, also with 31 students, was assigned to learn with the touchscreen-based concept mapping approach, while the control group, made up of 30 students, was arranged to learn with the paper-and-pencil-based concept mapping approach.

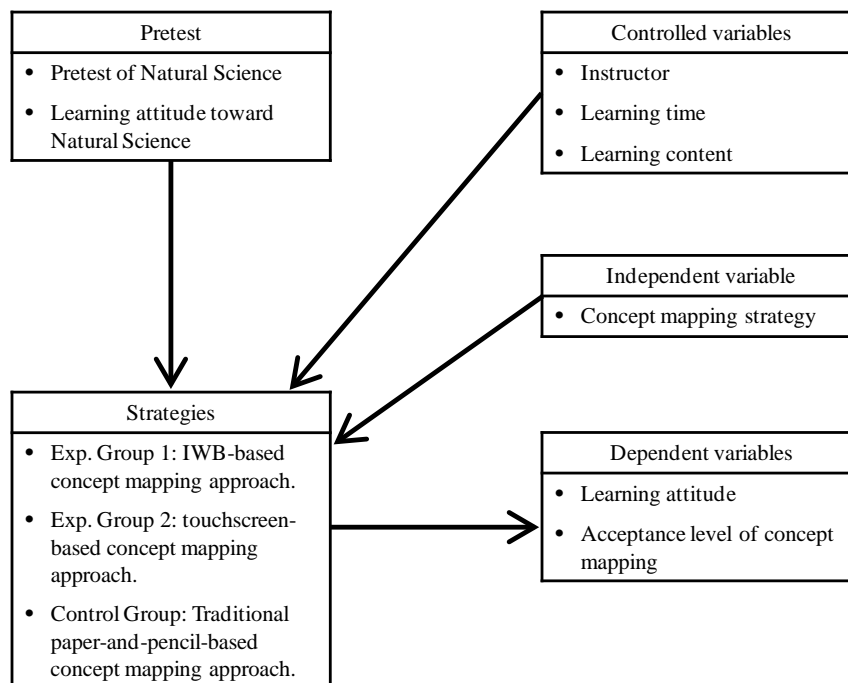


Figure 1. Research structure of the study

### Measuring tools

The measuring tools utilized in this study included a scale for measuring the students' attitudes toward the natural science course, and a questionnaire of the students' acceptance of the concept mapping approach.

The learning attitude questionnaire was adapted from the learning attitude scale developed by Crawley and Koballa Jr. (1994). It consisted of twenty-five items (as shown in the appendix) with a five-point Likert rating scheme, where 5 represented "strongly agree" and 1 represented "strongly disagree." The Cronbach's  $\alpha$  of this questionnaire was 0.862, showing high internal consistency and reliability (Cohen, 1988).

The acceptance questionnaire of concept mapping originated from the questionnaire developed by Chu, Hwang, Tsai, and Tseng (2010). It consisted of 31 items with a five-point Likert rating scheme, including 14 items for "Perceived usefulness" and 17 items for "Perceived ease of use." The Cronbach's alpha scores of the two dimensions were 0.90 and 0.91, respectively, showing the high internal consistency and reliability of the scale (Cohen, 1988).

### Experimental procedure

The experimental procedure of the study is depicted in Figure 2. The students were also asked to complete the learning attitude questionnaire prior to the activity.

During the learning activity, the students in the two experimental groups were given an introduction to the computerized concept mapping tools, while the control group was guided to learn the concept mapping development procedure with paper and pencil. Following the introduction to the concept mapping tool/development procedure, the three groups learned with the different concept mapping approaches.

The students in Experimental Group One were instructed by the teacher via the IWB-based concept mapping approach; that is, concept maps were mainly used as an interactive instructional tool via an IWB. The students in this group were asked to practice following the teacher's demonstration during the instruction process. The teacher first presented the concept mapping tool and some incomplete concept maps to the students on the IWB and asked them

to consider how to complete the maps. The teacher then asked some volunteers to come forward to fill in the missing parts (concepts or relationships between concepts) on the IWB. A discussion session was then conducted to allow all of the students to raise their hands and provide different opinions about the answers; that is, other students could also come forward to modify the concept maps on the IWB and explain why the modifications were needed. Finally, the teacher gave feedback to the students by showing the completed concept maps, and answered the questions raised by the students. Such an approach to conducting IWB-based instruction is commonly used in elementary schools in Taiwan.

The students in Experimental Group Two learned with the touchscreen-based concept mapping approach; that is, concept maps were used as a knowledge construction tool via computers equipped with touchscreens. The teachers first presented the concept mapping tool and the learning tasks, and then asked the students to develop their own concept maps using the tool. After the concept map development process, the teacher gave feedback to the students by showing the completed concept maps and answering the questions raised by the students.

In contrast to the two experimental groups, the students in the control group were guided to learn with the traditional concept mapping approach; that is, concept maps were used as a knowledge construction tool via paper and pencil. After the concept map development process, the teacher gave feedback to the students by showing the completed concept maps and answering the questions raised by the students.

After the learning activity, all of the students were asked to complete the questionnaires of learning attitude toward science learning and acceptance of using concept maps.

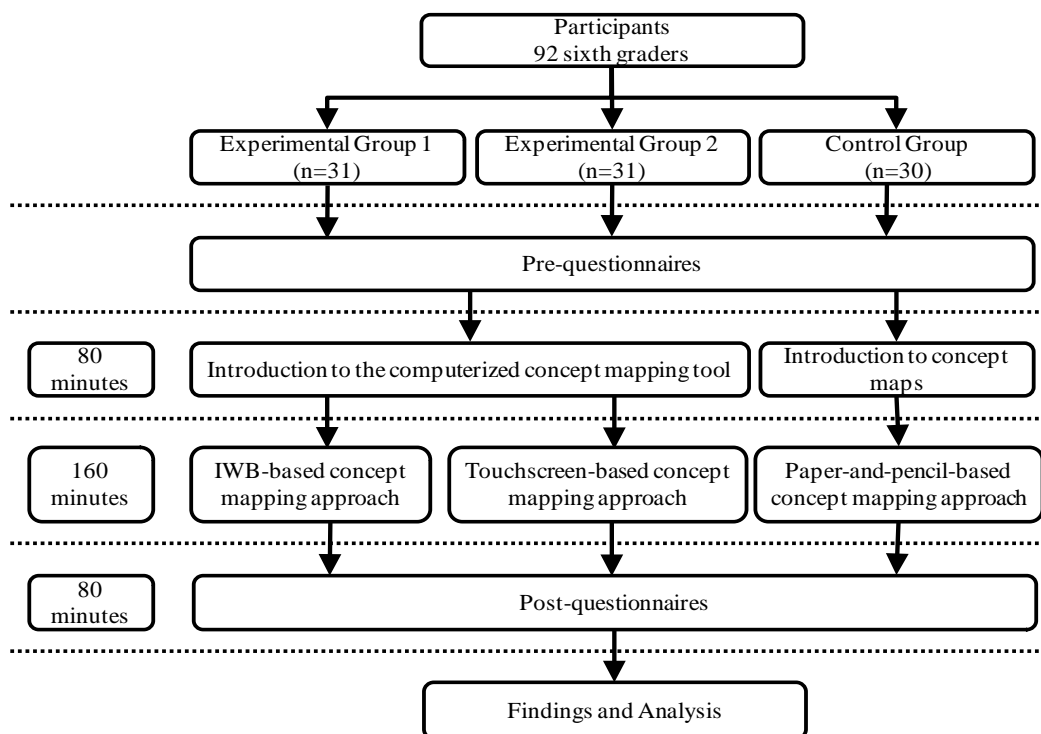


Figure 2. Experimental procedure of the study

## Results

### Analysis of learning attitudes toward the natural science course

To explore the effects of the different concept mapping approaches on the students' learning attitudes toward the natural science course, a paired-sample t-test was employed to compare their attitudes before and after the learning activity, as shown in Table 1.



It is found that the learning attitudes of the students in Experimental Group One showed significant improvement after participating in the learning activity ( $t = -2.88, p < .01$ ), indicating that the IWB-based instructional mode could promote the students' learning attitude; in contrast, the students in Experimental Group Two showed no significant difference in their attitudes after the learning activity ( $t = 0.70, p > .05$ ), indicating that the touchscreen-based learning mode had no significant effect on their learning attitudes towards the natural science course.

The analysis result of the control group was quite different from that of the two experimental groups as it actually showed a significant deterioration in attitudes after the learning activity ( $t = 3.34, p < .01$ ), indicating that the paper-and-pencil-based concept mapping mode could have significantly negative effects on students' learning attitudes.

*Table 1. Paired sample t-test of students' learning attitudes toward the natural science course for the three groups*

Group	Test	N	Mean	SD	<i>t</i>
Exp. Group 1	Pre-test	31	3.29	0.62	-2.88**
	Post-test	31	3.47	0.66	
Exp. Group 2	Pre-test	31	3.76	0.54	0.70
	Post-test	31	3.69	0.63	
Control Group	Pre-test	30	3.39	0.61	3.34**
	Post-test	30	3.05	0.57	

\*\*  $p < .01$

### Analysis of acceptance of concept mapping in the natural science course

To further explore the students' perceptions of using the different concept mapping approaches to learn natural science, one-way ANOVA was used to compare the students' ratings on the acceptance questionnaire items from two angles, that is, perceived usefulness and perceived ease of use, as shown in Table 2. The ANOVA result shows that Experimental Groups One and Two had significantly more positive perceptions of the usefulness of concept maps for learning than the control group ( $F = 6.37, p < .01$ ), indicating that the students highly accepted the computerized concept mapping approaches in terms of perceived usefulness, while the paper-and-pencil-based concept mapping strategy was perceived as being less effective.

*Table 2. ANOVA of perceived usefulness of concept mapping among the three groups*

Group	N	Mean	SD	<i>F</i>	Post hoc
(1) Exp. Group 1	31	3.70	0.67	6.37**	(1)>(3)
(2) Exp. Group 2	31	3.94	0.66		(2)>(3)
(3) Control Group	30	3.36	0.59		

\*\*  $p < .01$

In terms of perceived ease of use, Table 3 shows the ANOVA results of the ratings of the relevant questionnaire items. It was found that the two experimental groups revealed significantly better ratings than the control group ( $F = 8.32, p < .01$ ), indicating that the computerized concept mapping systems were perceived by the students as being much easier to use in comparison with the paper-and-pencil approach.

*Table 3. ANOVA for perceived ease of use of the three concept mapping approaches*

Group	N	Mean	SD	<i>F</i>	Post hoc
(1) Exp. Group 1	31	3.41	0.60	8.32**	(1)>(3)
(2) Exp. Group 2	31	3.69	0.62		(2)>(3)
(3) Control Group	30	3.04	0.67		

\*\*  $p < .01$

## Discussion

Although several previous studies have reported the effectiveness of the concept mapping approach, it remains an open issue to investigate the effects of using touch technologies on students' learning attitudes and technology

acceptance degrees in concept mapping activities. In this study, an experiment was conducted to compare students' learning attitudes and degrees of technology acceptance when using different types of touch technology-based concept mapping approaches. It was found that the students who learned with both the IWB and the touch screen-based concept mapping approaches revealed significantly better perceptions in terms of their learning attitudes toward the science course and the usefulness and ease of use of the concept mapping approaches. The findings of this study conform to those of some previous studies that applying novel technologies in learning activities has the potential to inspire students' extrinsic motivation and enhance their learning outcomes (Mistler-Jackson & Songer, 2000; Wang & Reeves, 2007). That is, the enjoyable experiences brought about by using touch technologies could be one of the factors that motivated the students. Moreover, the use of touch technology enabled the students to easily edit and directly operate on their learning tasks, which could be the factor affecting their perceived ease of use and usefulness owing to a reduction in the Split-Attention Effect indicated by Mayer and Moreno (1998, 2003).

Moreover, it is interesting to find that the average learning attitude rating of the students in Experimental Group Two decreased after the learning activity, although the decrease was not significant. Moreover, the average rating of the control group was significantly decreased. From the interviews, it was found that the students in Experimental Group Two and the control group showed less interest in the learning activity since they spent most of their time developing concept maps (i.e., their learning tasks), while the students in Experimental Group One enjoyed the interactions with the teacher and their peers via the IWB. This finding implies the need to design some interaction or competitive activities in the concept map development tasks.

It should also be noted that the use of touch screens by students is more learner-centered, while the interactive whiteboard approach is more teacher-centered; however, this does not imply that the findings are trivial since student-centered approaches are not always more effective than teacher-centered approaches, in particular for elementary school students who might require more instruction or guidance. In this study, although the teacher attempted to interact with students using the IWB and to guide them in discussion, the students had to mainly listen to the teacher and exchange information; thus, they had less time to organize their knowledge. However, from the learning attitude ratings of the students, it can be seen that the IWB-based approach has the potential to promote students' learning interest and hence improve their learning attitudes. Consequently, in the future, more thoughtful learning design of IWB-based instruction could take into account the time required for knowledge organization.

To investigate the factors affecting the performance of the students in Experimental Group One, an in-depth interview was arranged. Seven students from the group were randomly selected to take part in the interview.

The interview results showed that most of the students in this group liked the IWB-based learning approach. Several students indicated that, in such an interesting interaction mode, the teacher presented the notation and illustrative examples of the learning content in the form of concept maps, and interacted with the students on the IWB by asking them to answer some questions or fill in some concepts or the relationships between concepts. However, several students also pointed out the problems of such a teacher-centered instructional mode of using concept maps, including "less time to practice" and "less opportunity to reflect." For example, one of the students stated that *"Although it was fun to interact with the teacher and my classmates on the IWB, I did not have much time to organize what I learned in this subject unit."* Another student stated that *"Lack of sufficient practice was the problem of learning in this way. I liked to interact with everyone, but there were dozens of classmates in the class and each of us had only a few chances to practice on the IWB. Most of the time we just watched the teacher's presentation and the interactions between the teacher and our classmates."*

It can be seen from the interview responses that the IWB-based concept mapping approach is able to attract the attention of students; however, with such an approach, the students were less involved since the concept maps were used as an instructional tool. On the contrary, each of the students in Experimental Group Two used a computer with a touchscreen to develop concept maps based on what they had learned in the class; that is, they learned with a student-centered approach that allowed them to have sufficient time to practice and organize their knowledge. Therefore, insufficient time to practice and organize knowledge could be one main reason why the learning achievement of the students who learned with the IWB-based concept mapping approach was worse than that of the students who learned with the touchscreen-based approach.

Although the learning approach used by the control group students was student-centered, the efficiency of using paper and pencils for developing concept maps was not as good as that of the touchscreen-based approach, because it

is inconvenient for the control group students to draw and modify concept maps with paper and pencil. Such a learning context could lead to a decrease in their learning efficiency, which could be a factor influencing their learning achievement. Previous studies have also found that the performance of students who learn with paper and pencil is usually inferior to that of those who learn with technologies, in terms of learning achievement and motivation (Komis, Ergazaki & Zogza, 2007; Kordaki, 2009; Liu, 2011). Therefore, the students' learning attitudes toward the natural science course appeared to be negative after the learning activity. Moreover, their feedback regarding the "perceived usefulness" and "perceived ease of use" was significantly worse than that of the other two groups. Such findings conform to what has been reported by previous studies, namely that computerized concept mapping approaches are more readily accepted by students in comparison with the traditional paper-and-pencil approach (Hwang, Shi & Chu, 2011; Trundle & Bell, 2010; Wu et al., 2012).

It should be noted that the three concept mapping approaches reflected the teaching reality of most schools in Taiwan. The experiment design aimed to investigate the effects of using touch technologies in practical applications on students' learning attitudes and perceptions in comparison with the traditional paper-and-pencil approach. Both the teaching methods applied to the touch technologies followed the real instructional designs in the selected school (and many other schools) as well as the nature of the technologies, although several factors could be involved and might not be easy to control. Consequently, in the future, it is worth investigating impacts of individual factors in depth by conducting extended experiments. Moreover, it is also worth evaluating the effectiveness of those touch technology-based concept mapping approaches from the aspects of human factors, such as genders, learning styles or cognitive styles (Hwang, Sung, Hung, & Huang, 2012).

## Conclusions

Many past studies have shown that concept mapping strategies could improve students' academic achievement (Erdogan, 2009; Hwang, Shi & Chu, 2011; Lim, Lee & Grabowski, 2009). What is more, researchers have also indicated that a potential way for educators to conduct concept map-based instruction is to use computerized tools in place of traditional paper-and-pencil-based approaches (Kim & Olaciregui, 2008). This study investigates the effects of two different touch technology-based concept mapping interaction modes on students' learning attitudes and learning achievement. The research findings of the study are summarized as follows:

- *The IWB-based concept mapping approach could improve students' learning attitudes toward the natural science course, while using concept maps with paper and pencil might have a detrimental effect on students' learning attitudes.*

The experimental results show that the learning attitudes of the students in Experimental Group One (i.e., the teacher-centered IWB-based approach) significantly improved after the learning activity, indicating that the IWB-based interaction mode promoted the students' learning attitudes towards the natural science course. On the contrary, the students in the control group (i.e., the traditional paper-and-pencil-based approach) revealed significantly worse attitudes after the learning activity. Such a finding conforms to some of the previous studies that report the differences between the learning motivations and attitudes of students using computerized and paper-and-pencil concept mapping approaches (Hwang, Shi, & Chu, 2011; Wu et al., 2012).

- *Both computerized approaches, including teacher-centered (i.e., IWB-based) and student-centered (i.e., touchscreen-based), were highly accepted by the students in terms of the perceived usefulness and perceived ease of use of the concept maps for learning.*

The experimental results concerning the students' degree of acceptance of using concept maps during the learning activity are encouraging. It can be seen that both forms of touch technology-based computerized concept mapping approaches were highly recognized by the students as being useful and easy to use. That is, with the help of the touch technologies (i.e., the IWB and computers with touchscreens), the students felt that learning with concept maps was not difficult and was helpful to them in understanding the learning content. Such a finding conforms to most of the previous studies that have recognized the effectiveness of computerized concept mapping strategies (Liu, Chen & Chang, 2010; Reader & Hammond, 1994).

From the findings and the above discussion, it is concluded that touch technology-based concept mapping approaches are highly accepted by students in terms of their perceived usefulness and ease of use; moreover, the teacher-centered IWB-based approach seems to be effective in improving the students' learning attitudes, while the student-centered touchscreen-based approach seems to be effective in improving their learning achievements. Consequently, it is inferred that the IWB-based approach is a good instructional strategy for teachers to adopt when interacting with students and to gain their attention, while the touchscreen-based approach provides students with a good interface for developing their concept maps. To sum up, it is suggested that teachers and educators note the importance and strategies of employing touch technologies in conducting concept mapping-related learning activities (Smith, Higgins, Wall, & Miller, 2005).

In the future, it is worth developing a touch technology-based concept mapping system which not only allows students to develop their own concept maps, but also allows teachers to interact with students during the learning activities, that is, a computerized concept mapping approach that integrates the teacher-centered IWB and the student-centered touchscreen technologies to benefit students in terms of improving both their learning attitudes and their learning achievements. It is also worth investigating the effects of other innovative touch technologies, such as multi-touch interfaces, on students' learning performance in concept mapping-related learning activities. Moreover, as numerous previous studies have employed various computerized concept mapping approaches, it would be interesting to compare the effect of touch technology-based concept mapping with other kinds of computerized concept mapping in future studies.

## Acknowledgements

The authors would like to thank Mr. Ze-You Wang, an elementary school teacher, for his assistance in conducting the experiment. This study is supported in part by the National Science Council of the Republic of China under contract numbers NSC 99-2511-S-011-011-MY3 and NSC 101-2511-S-011 -005 -MY3.

## References

- Ausubel, D. P. (1963). *The psychology of meaningful verbal learning*. New York, NY: Grune & Stratton.
- Chu, H. C., Hwang, G. J., Tsai, C. C., & Tseng, J. C. R. (2010). A two-tier test approach to developing location-aware mobile learning systems for natural science course. *Computers & Education*, 55(4), 1618-1627.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Crawley, F. E., & Koballa Jr, T. R. (1994). Attitude research in science education: Contemporary models and methods. *Science Education*, 78(1), 35-55.
- Dhindsa, H. S., & Emran, S. H. (2006). Use of the interactive whiteboard in constructivist teaching for higher student achievement. In S. M. Stewart, J. E. Olearski, & D. Thompson (Eds), *Proceedings of the Second Annual Conference for Middle East Teachers of Science, Mathematics and Computing* (pp. 175-188). Abu Dhabi, United Arab Emirates: METSMaC.
- Erdogan, Y. (2009). Paper-based and computer-based concept mappings: The effects on computer achievement, computer anxiety and computer attitude. *British Journal of Educational Technology*, 40(5), 821-836.
- Goodison, T. A. (2002). Learning with ICT at primary level: Pupils' perceptions. *Journal of Computer Assisted Learning*, 18(3), 282-295.
- Hollan, J. D., & Hutchins, E. L. (2010). Opportunities and challenges for augmented environments: A distributed cognition perspective. In S. Lahlou (Ed.), *Designing User Friendly Augmented Work Environments: From Meeting Rooms to Digital Collaborative Spaces* (pp. 237-259). doi: 10.1007/978-1-84800-098-8\_9
- Hung, P. H., Hwang, G. J., Su, I. H., & Lin, I. H. (2012). A concept-map integrated dynamic assessment system for improving ecology observation competences in mobile learning activities. *Turkish Online Journal of Educational Technology*, 11(1), 10-19.
- Hung, P. H., Lin, Y. F., & Hwang, G. J. (2010). Formative assessment design for PDA integrated ecology observation. *Educational Technology & Society*, 13(3), 33-42.
- Hwang, G. J. (2003). A conceptual map model for developing intelligent tutoring systems. *Computers & Education*, 40(3), 217-235.

- Hwang, G. J., Shi, Y. R., & Chu, H. C. (2011). A concept map approach to developing collaborative Mindtools for context-aware ubiquitous learning. *British Journal of Educational Technology*, 42(5), 778-789.
- Hwang, G. J., Sung, H. Y., Hung, C. M., & Huang, I. (2012). Development of a personalized educational computer game based on students' learning styles. *Educational Technology Research & Development*, 60(4), 623-638.
- Hwang, G. J., Tsai, C. C., Chu, H. C., Kinshuk, & Chen, C. Y. (2012). A context-aware ubiquitous learning approach to conducting scientific inquiry activities in a science park. *Australasian Journal of Educational Technology*, 28(5), 931-947.
- Hwang, G. J., Wu, P. H., & Ke, H. R. (2011). An interactive concept map approach to supporting mobile learning activities for natural science courses. *Computers & Education*, 57(4), 2272-2280.
- Kennewell, S. (2005, September). *Researching the influence of interactive presentation tools on teacher pedagogy*. Paper presented at the British Educational Research Association Annual Conference, University of Glamorgan, Wales, UK.
- Khamesan, A., & Hammond, N. (2004). Interpersonal awareness during web-based concept mapping: The effect of different communication channels. In L. Cantoni & C. McLoughlin (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2004* (pp. 4168-4173). Chesapeake, VA: AACE.
- Kim, P., & Olaciregui, C. (2008). The effects of a concept map-based information display in an electronic portfolio system on information processing and retention in a fifth-grade science class covering the Earth's atmosphere. *British Journal of Educational Technology*, 39(4), 700-714.
- Komis, V., Ergazaki, M., & Zogza, V. (2007). Comparing computer-supported dynamic modeling and 'paper & pencil' concept mapping technique in students' collaborative activity. *Computers & Education*, 49(4), 991-1017.
- Kordaki, M. (2009). A drawing and multi-representational computer environment for beginners' learning of programming using C: Design and pilot formative evaluation. *Computers & Education*, 54(1), 69-87.
- Kwon, S. Y., & Cifuentes, L. (2009). The comparative effect of individually-constructed vs. Collaboratively-constructed computer-based concept maps. *Computers & Education*, 52(2), 365-375.
- Lim, K. Y., Lee, H. W., & Grabowski, B. (2009). Does concept-mapping strategy work for everyone? The levels of generativity and learners' self-regulated learning skills. *British Journal of Educational Technology*, 40(4), 606-618.
- Liu, P. (2011). A study on the use of computerized concept mapping to assist ESL learners' writing. *Computers & Education*, 57(4), 2548-2558.
- Liu, P. L., Chen, C. J., & Chang, Y. J. (2010). Effects of a computer-assisted concept mapping learning strategy on EFL college students' English reading comprehension. *Computers & Education*, 54(2), 436-445.
- Morgan, M., & Butler, M. (2009). *Considering multi-touch display technology for collaboration in the classroom*. In G. Siemens & C. Fulford (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2009* (pp.674-683). Chesapeake, VA: AACE.
- Markham, K. M., Mintzes, J. J., & Jones, M. G. (1994). The concept as a research and evaluation tool: Further evidence of validity. *Journal of Research in Science Teaching*, 31(1), 91-101.
- Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, 90, 312-320.
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38(1), 43-52.
- McClure, J. R., Sonak, B., & Suen, H. K. (1999). Concept map assessment of classroom learning: Reliability, validity, and logistical practicality. *Journal of Research in Science Teaching*, 36(4), 475-492.
- Mistler-Jackson, M., & Songer, N. B. (2000). Student motivation and Internet technology: Are students empowered to learn science? *Journal of Research in Science Teaching*, 37(5), 459-479.
- Novak, J. D. (1990). Concept mapping: A useful tool for science education. *Journal of Research in Science Teaching*, 27(10), 937-949.
- Novak, J. D. (1996). *Concept mapping: A tool for improving science teaching and learning*. In D. F. Treagust, R. Duit, & B. J. Fraser(Eds.), *Improving Teaching and Learning in Science and Mathematics* (pp.32-43). New York, NY: Teachers College Press.
- Novak, J. D., & Cañas, A. J. (2006). The origins of the concept mapping tool and the continuing evolution of the tool. *Information Visualization Journal*, 5(3), 175-184.

- Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. Cambridge, UK: Cambridge University press.
- Novak, J. D., Gowin, D. B., & Johansen, G. T. (1983). The use of concept mapping and knowledge vee mapping with junior high school science students. *Science Education*, 67(5), 625-645.
- Panjaburee, P., Hwang, G. J., Triampo, W., & Shih, B. Y. (2010). A multi-expert approach for developing testing and diagnostic systems based on the concept effect model. *Computers & Education*, 55(2), 527-540.
- Reader, W., & Hammond, N. (1994). Computer-based tools to support learning from hypertext: concept mapping tools and beyond. *Computers & Education*, 12(1), 99-106.
- Rick, J., & Rogers, Y. (2008, October). From digiquilt to digitile: Adapting educational technology to a multi-touch table. Paper presented at the 3rd IEEE International Workshop on Horizontal Interactive Human Computer Systems, Amsterdam, Netherlands.
- Romeo, G., Edwards, S., McNamara, S., Walker, I., & Ziguras, C. (2003). Touching the screen: issues related to the use of touchscreen technology in early childhood education. *British Journal of Educational Technology*, 34(3), 329-339.
- Roth, W. M., & Roychoudhury, A. (1994). Science discourse through collaborative concept mapping: New perspectives for the teacher. *International Journal of Science Education*, 16(4), 437-455.
- Ruiz-Primo, M. A., Shavelson, R. J., & Schultz, S. E. (1997). *On the validity of concept map-based assessment interpretations: An experiment testing the assumption of hierarchical concept maps in science*. (Center for the Study of Evaluation (CSE) Report No. 445). Los Angeles, CA: Graduate School of Education & Information Studies University of California.
- Schmid, R. F., & Telaro, G. (1990). Concept mapping as an instructional strategy for high school biology. *Journal of Education Research*, 84(2), 78-85.
- Smith, A. (2000). *Interactive whiteboard evaluation*. Retrieved May 31, 2012, from the MirandaNet Publications website: <http://www.mirandanet.ac.uk/pubs/smartboard.htm>
- Smith, H. J., Higgins, S., Wall, K., & Miller, J. (2005). Interactive whiteboard: Boon or bandwagon? A critical review of the Literature. *Journal of Computer Assisted Learning*, 21(2), 91-101.
- Somekh, B., Haldane, M., Jones, K., Lewin, C., Steadman, S., Scrimshaw, P., ... Woodrow, D. (2007). *Evaluation of the primary schools whiteboard expansion project* (Report to the Department for Children, Schools and Families). Manchester, UK: Education & Social Research Institute, Manchester Metropolitan University.
- Trundle, K. C., & Bell, R. L. (2010). The use of a computer simulation to promote conceptual change: A quasi-experimental study. *Computers & Education*, 54(4), 1078-1088.
- Wang, S. K. & Reeves, T. C. (2007). The effects of a web-based learning environment on student motivation in a high school earth course. *Educational Technology, Research and Development*, 55(2), 169-192.
- Wood, R., & Ashfield, J. (2008). The use of the interactive whiteboard for creative teaching and learning in literacy and mathematics: A case study. *British Journal of Educational Technology*, 39(1), 84-96.
- Wu, P. H., Hwang, G. J., Milrad, M., Ke, H. R., & Huang, Y. M. (2012). An innovative concept map approach for improving students' learning performance with an instant feedback mechanism. *British Journal of Educational Technology*, 43(2), 217-232.
- Wu, P. H., Hwang, G. J., Su, L. H., & Huang, Y. M. (2012). A context-aware mobile learning system for supporting cognitive apprenticeships in nursing skills training. *Educational Technology & Society*, 15(1), 223-236.
- Xu, H. L., & Moloney, R. (2011). Perceptions of interactive whiteboard pedagogy in the teaching of Chinese language. *Australasian Journal of Educational Technology*, 27(2), 307-325.
- Yu, X., Zhang, M., Ren, J., Zhao, H., & Zhu, Z. (2010). Experimental development of competitive digital educational games on multi-touch screen for young children. *Lecture Notes in Computer Science*, 6249, 367-375.

## Appendix

### Questionnaire of students' learning attitudes toward the natural science course

(\*inverse items)

1. The learning content of the natural science course is plentiful and helpful.
2. The learning content of the natural science course can benefit me a lot.
3. I have fully concentrated on the learning activity.
4. I often do my own thing in the natural science class.\*
5. I am very interested in the natural science class.
6. The learning content of the natural science course is not attractive to me.\*
7. The learning content of the natural science course is useful.
8. I can gain a lot from the natural science course.
9. I don't like the natural science course.\*
10. The knowledge learned from the natural science course is useless in daily life.\*
11. I like to do the assignments of the natural science course.
12. It is important to actively solve problems encountered in learning natural science.
13. I expect to gain good grades in the natural science course.
14. I shall apply what I have learned in the natural science course to daily life.
15. I hate the course since it often needs me to think and make reflections.\*
16. I like to propose questions and discuss with the natural science teacher.
17. I expect to learn to solve problems by myself in the natural science course.
18. It is important to spend time on previewing and reviewing the natural science course content at home.
19. The learning content of the natural science course is helpful to me in observing the natural phenomena of daily life in depth.
20. The learning content of the natural science course is important.
21. It is important to do the assignments and prepare for examinations of the course on time.
22. I like to discuss with teachers and peers when I cannot fully understand the learning content.
23. It is important to take notes on the key points of the learning unit.
24. I usually give up when encountering problems during the learning process.\*
25. I like to take assessments to assure that I really understand the learning content.

## Learning by Searching: A Learning Environment that Provides Searching and Analysis Facilities for Supporting Trend Analysis Activities

Chengjiu Yin<sup>1</sup>, Han-Yu Sung<sup>2</sup>, Gwo-Jen Hwang<sup>3\*</sup>, Sachio Hirokawa<sup>1</sup>, Hui-Chun Chu<sup>4</sup>,  
Brendan Flanagan<sup>5</sup> and Yoshiyuki Tabata<sup>1</sup>

<sup>1</sup>Research Institute for Information Technology, Kyushu University, Fukuoka, Japan // <sup>2</sup>Graduate Institute of Applied Science and Technology, National Taiwan University of Science and Technology, Taipei, Taiwan // <sup>3</sup>Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taipei, Taiwan // <sup>4</sup>Department of Computer Science and Information Management, Soochow University, Taipei, Taiwan // <sup>5</sup>Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka, Japan // yin@cc.kyushu-u.ac.jp // hanyu.sung@gmail.com // gjhwang.academic@gmail.com // hirokawa@cc.kyushu-u.ac.jp // carolhcchu@gmail.com // bflanagan.kyudai@gmail.com // tabata@cc.kyushu-u.ac.jp \* corresponding author

(Submitted February 12, 2012; Revised September 08, 2012; Accepted September 24, 2012)

### ABSTRACT

With the popularity of the Internet, online searching is becoming an important part of learning. In this paper, based on the “Learning by Searching” theory, a learning environment is developed, which includes a search engine to assist students in recognizing the progression of trends and keyword transitions for specific domains. To efficiently support research trend surveys, an automatic data accumulation and classification approach is proposed to construct the database excerpts instead of manual keyword registration or any other heuristic preprocesses. With an associative search module, the search engine dynamically searches for relevant words that are frequently used in the targeted academic field, and provides learners with effective visualizations to understand the trend transitions. An experiment has been conducted on a college information management course to show the effectiveness of the proposed approach. The experiment results show that the students who learned with the new approach had significantly better learning performance in terms of recognizing the trend transitions of the targeted issues than those who learned with conventional search engines.

### Keywords

Learning by searching, Web-based learning, Search engine, Data mining, Research trend survey

### Background and objectives

With the advancement and popularity of the Internet and search engine technology, fostering students’ web-based information searching ability has become an important educational objective. Researchers have indicated that allowing students to learn as active and self-directed participants is one of the greatest benefits of web-based learning activities, which often involve information searching tasks (Bilal, 2000; Hwang, Tsai, Tsai, & Tseng, 2008). In the past decade, various issues concerning information-seeking have been studied, such as the skill of processing the searched information (Chiou, Hwang, & Tseng, 2009) and the development of new environments that facilitate teachers’ observation and analysis of the information-seeking behaviors of students in web-based learning environments (Tseng, Hwang, Tsai, & Tsai, 2009). Moreover, the correlates of teachers’ epistemological beliefs concerning Internet environments, online search strategies, and search outcomes have been investigated (Tsai, Tsai, & Hwang, 2011).

In the meantime, researchers have indicated the difficulty of fostering students’ higher order thinking competences, such as “Evaluate” and “Analyze” in Bloom’s taxonomy of educational objectives (Bloom et al., 1956; Anderson et al., 2001; Hwang, Chu, Lin, & Tsai, 2011). Although the existing search engines cater to students’ basic knowledge acquisition, they are not categorized into special research areas, making it difficult to address the specific, unique needs of individual learners. The question is how to design better search engines that address users’ learning needs and knowledge levels. An ideal search engine should not only show the retrieval results, but also the analysis. Fortunately, technologies can accelerate learning and boost creativity. With the development of technologies such as data-processing, it is possible to design better search engines to address learning needs. Data-processing includes functions such as search engines, data mining, recommendations, and so on.



Moreover, with the development of technology, the longevity of paper-based literature has become very short. Researchers need to update their knowledge constantly in order to keep up with technological advancements. Hwang and Tsai (2011) indicated that "analysis results could help policymakers in governments and researchers in professional organizations to allocate the necessary resources and make plans for supporting future research and applications." They also indicated that doing surveys could provide good references for educators and researchers who plan to contribute to the relevant studies. Therefore, it is essential to conduct surveys to have both a wide and a deep understanding of related research; in particular, for those students who are just beginning to engage in academic research, research trend surveying is an essential preliminary step for any academic research (Hwang & Wu, 2012; Karatas, 2008). Doing an academic research survey can foster their competences of collecting the information related to a specific topic, evaluating the collected data, and analyzing the trends in the field. Unfortunately, although students can gain knowledge by using the current search engines such as Google or Yahoo!, many novice researchers have difficulty analyzing the collected data without proper support from the learning environment. Therefore, it is necessary to create search engines which are dedicated to supporting the acquisition of knowledge according to the special research area, to effectively supporting "learning by searching".

In this paper, a learning system, "Milky Way Research Trend (MWRT)," that includes a search engine for supporting research trend surveys for scientific literature is proposed. The search engine provides students with an efficient literature survey tool, which not only allows access to the needed data, but also presents the analysis results of the trends. Instead of manual keyword registration or any other heuristic preprocesses, the search engine dynamically searches for relevant words that are frequently used in the targeted academic field, and presents the findings in a visual format to help students understand the trend transitions. With this system, students can perform trend analyses, automatically extract outlines from the literature, and analyze the targeted documents as a time-series.

To evaluate the effectiveness of the proposed approach, an experiment has been conducted in a college information management course to investigate the following research questions:

1. Do the students who learn with MWRT show better learning outcomes in trend analysis than those who learn with the conventional learning approach for trend analysis?
2. Are there significant differences between the cognitive loads of the students who learn with MWRT and those who learn with the conventional learning approach for trend analysis?
3. Do the students who learn with MWRT reveal different technology acceptance degrees than those who learn with the conventional learning approach in terms of perceived ease of use and perceived usefulness?

## **Learning by searching**

### **Searching is a natural learning behavior**

Searching is a natural learning behavior like listening, speaking, reading or writing. There are many reasons for people to seek information. Sometimes people search for information because of curiosity; that is, they want to know why. Sometimes they search for information purely for their need to solve problems or complete tasks. Whatever the actual reason, the information searching process is a cognitive process that acquires knowledge actively, which is defined as a way of learning referred to as "learning by searching" in this study.

There are many kinds of learning strategies, such as learning by attending classes, learning from informal incidents, learning by doing, learning by gaming, and learning by searching. Among these learning strategies, learning by searching can foster students' ability to take the initiative to acquire knowledge (Hwang, Tsai, Tsai, & Tseng, 2008). This research advocates learning by searching. It is a method for promoting "discovery learning," which is an inquiry-based, constructivist learning theory. Discovery learning takes place in problem solving situations where the learner draws on his or her own past experience and existing knowledge to discover facts and relationships and new truths to be learned (Bruner, 1967). Students interact with the world by exploring and manipulating objects, wrestling with questions and controversies, or performing experiments. Bruner suggested that students are more likely to remember concepts if they discover them on their own. Several studies have reported the importance of engaging students in discovery learning tasks using search engines to seek information. For example, Bilal (2000) pointed out the importance and necessity of engaging students in searching for information on the Internet; in the meantime, he also indicated that, for those students who are unfamiliar with the usage of search engines, incorrect or inappropriate

information is likely to be derived. Accordingly, Hwang and Kuo (2011) proposed an information-summarizing instruction strategy to help students improve their information-searching ability. They found that, with proper supports, the students' ability of using keywords, selecting information resources and extracting important content can be improved. Kuo, Hwang and Lee (2012) further defined the learning approach that employs search engines to search for information to answer a series of questions related to a specific topic as "web-based problem solving."

Most of the studies concerning web-based information seeking employ the existing search engines to support discovery learning or web-based problem solving. Such an approach might be helpful to students in collecting data to complete reports for specific learning tasks, or finding information to answer a series of questions. However, to support students in investigating the trends of specific research topics or in analyzing the shifts of technology advancements and applications, more effective tools are needed. In this study, an innovative search engine that helps students organize the searched knowledge in order to improve their trend-awareness ability is proposed to cope with this problem.

### Categories of knowledge

Searching can be perceived as a process of acquiring knowledge. When seeking knowledge on the Internet, people often represent their quests with one of the following "5W1H" questions (Tseng & Hwang, 2007); that is, "What", "Where", "Which", "Who", "When" and "How". Researchers have indicated that knowledge can be divided into two categories; one is ability, which includes "know how", "know what", "know when" and "know who"; the other is related to knowing where to find the knowledge needed, which includes "know where" and "know which". As shown in Figure 1, ability is supplemented with knowing where. As knowledge continues to grow and evolve at a rapid pace, access to what is needed is more important than what the learner currently possesses (Siemens, 2004).

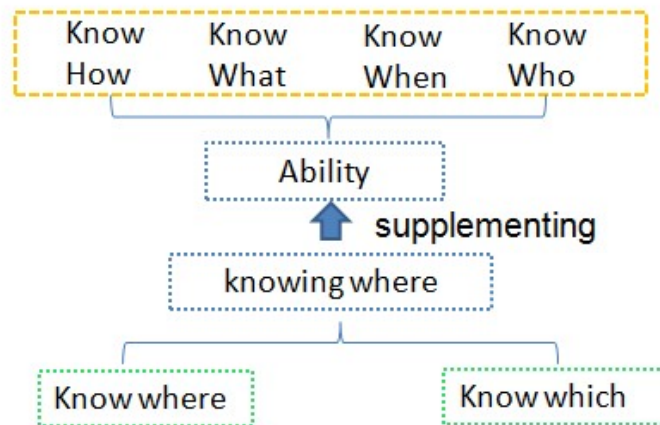


Figure 1. Knowledge Relationships

With the advancement and popularity of the Internet and search engine technology, online information searching is recognized as being a part of learning and a required learning skill for students (Liu, 2008). Researchers have attempted to investigate the cognitive processes underlying information searching. For example, State (2009) examined the search habits of 72 participants while conducting a total of 426 searching tasks. It was found that search engines were mainly used for checking the learners' own internal knowledge via comparing the knowledge with the searched facts, meaning that information searching is a learning process rather than simply a way of obtaining information.

This study aims to develop a learning system with a search engine which is able to advise students how to organize the collected trend information as well as how to find the information for solving practical problems. Via using the learning system, the students learn to recognize the research trends in depth through selecting keywords, setting search parameters and interpreting the search results.

## System description

The data processing procedure consists of three stages; that is, “data collection,” “development of the search engine,” and “data analysis using the search engine.”

### Data collection

The MWRT search engine was populated with 13,326 articles and papers from SciVerse Scopus (<http://www.info.sciverse.com/scopus/>) that were published from the year 1992 to 2012, and which contained the keyword “e-learning”. This data set was used in the analysis conducted in this paper. The original “SciVerse Scopus” database system does not provide any tools for high level analysis of the documents obtained. Therefore, a special search engine was developed for performing high level analysis of the documents collected.

### Development of the search engine

The learning system was developed on an Apache server using the Linux operating system and the Perl programming language. The learner enters the keywords about his/her research field to search for relevant data. The relevance of inputted keywords is calculated by our system using co-occurrence frequency. This means that if a word has a higher relevance to the keywords given in the abstracts of papers (i.e., co-occurrence frequency), these words are listed as feature words. Subsequently, the system gives the user effective visualizations to understand the research trend transitions.

### Data analysis using the search engine

The MWRT system provides an efficient tool to assist students in doing research trend analysis. Using the MWRT search engine, students can perform trend analyses, automatically extract outlines from the literature, and analyze documents as a time-series. Figure 2 shows the interface of the system. The user can control the MWRT search engine by changing certain inputs and parameters (a) as displayed in Figure 2. Based on these, the results (b) are then output as displayed in Figure 3.

#### (a) Input query and parameters

1: Search query string; 2: Enable advanced search settings display; 3: Enable detailed results output; 4: Number of top ranked keywords for each year; 5: Number of top ranked keys for each year; 6: Results sorting method; 7: Number of top ranked countries for each year; 8: Year range; 9: Number of top ranked authors for each year; 10: Threshold of the minimum number of top ranking years; 11: Number of top ranked sources for each year; 12: Toggle to display results as dots instead of numeric output; 13: Number of top ranked organizations for each year.

The screenshot shows the input interface of the MWRT search engine. It includes a search bar with the text "mobile learning", a "Search" button, and several checkboxes for "detail Menu", "detail List", "stop", "dot", and "debug". Below the search bar, there are various input fields and dropdown menus for parameters: "top words by" (set to "weight"), "from" (1992), "to" (2012), "T >=" (1), "word" (5), "key", "country", "author", "source", and "organization". Arrows point from numbered labels (1-13) to specific elements in the interface: 1 points to the search bar, 2 points to the "detail Menu" checkbox, 3 points to the "detail List" checkbox, 4 points to the "word" dropdown, 5 points to the "key" dropdown, 6 points to the "country" dropdown, 7 points to the "author" dropdown, 8 points to the "source" dropdown, 9 points to the "organization" dropdown, 10 points to the "T >=" dropdown, 11 points to the "stop" checkbox, 12 points to the "dot" checkbox, and 13 points to the "debug" checkbox.

Figure 2. Input query and parameters interface

#### (b) Results output

1: Total number of articles containing the query string; 2: The number of years the item was top ranked; 3: The number of articles that contained the item in the top ranked years; 4: The total number of articles that contain the item; 5: The Milky Way which displays the top ranking years for each item.

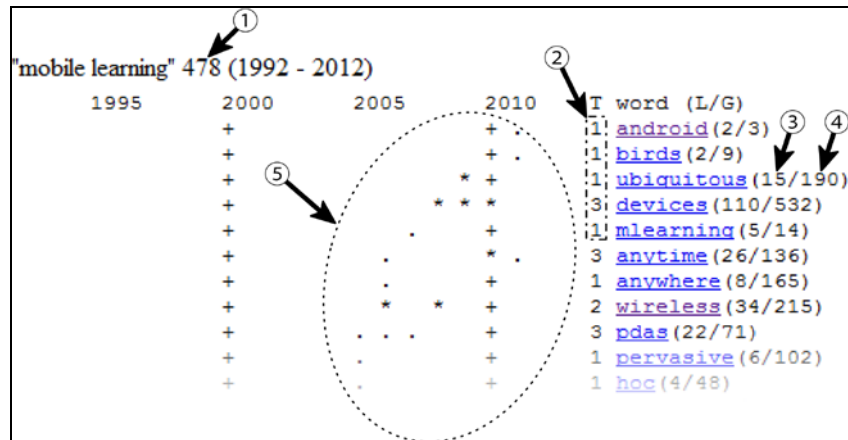


Figure 3. Results output interface

## Analysis examples

To demonstrate the effectiveness of the system, several practical applications of analyzing research trends with MWRT are given in this section.

### Research trends overview by feature words

In the following, an illustrative example of investigating research trends of mobile learning is given to demonstrate the "Overview by Feature Words" function of the learning system. In this case, the analysis was conducted using the keywords "mobile learning" in the MWRT search engine and the number of top ranked keywords to be returned for each year was set to five.

Some definitions of "mobile learning" describe it as the act of learning while on the move. In more recent times the definition has become associated with the use of mobile computing devices used to facilitate the act of learning (Laouris & Eteokleous, 2005). It is hypothesized that this has been spurred by advances in mobile communication devices, such as laptops, mobile phones, smart phones and more recently tablet computers (Ha et al., 2009). The MWRT search engine was used to analyze the research trends in mobile learning to test this hypothesis.

The results of the analysis are displayed in Figure 4. As observed in 2005, three technology-related keywords emerged as possible research trends, "phones", "pdas" and to a lesser extent, "hoc". On further investigation into the usage of these keyword it was determined that "pdas" referred to the use of Personal Digital Assistants or Pocket PCs, while "phones" referred mainly to the use of mobile (cellular) phones and Smartphones used to facilitate learning in a mobile environment. The keyword "hoc" was found to refer to "ad hoc" wireless communication networks. These keywords focus on hardware based technologies that are used to support mobile learning.

The keyword "wireless" comes to light in 2006 with additional research conducted in wireless Internet and network communications, as well as research into the use of near field wireless communication, such as Bluetooth personal area networks and infrared data communications. In 2008 the keyword "devices" appears as a research trend. On further investigation it was found to refer to the use of Smartphones, Pocket PCs, PDAs, mobile phones, portable computers (such as laptops and notebooks) and to a lesser extent, mobile music players, and digital dictionaries. Up until this point hardware based technologies were dominant in the research trends analysis results.

The most recent keyword, "android", first surfaced in 2011 as a reference to the Android Smartphone operating system software by Google which has grown in consumer market popularity in recent years since its release in 2008 (The Nielsen Company, 2011). This change in the trend from hardware to software based technologies may suggest that future research trends will focus on emerging mobile software platforms as opposed to previously dominant hardware devices.

This trail of keywords returned in the results from the MWRT search engine suggests that a technology driven research trend in mobile learning exists and offers possible insight for novice research students into the change from hardware to software technology based research.

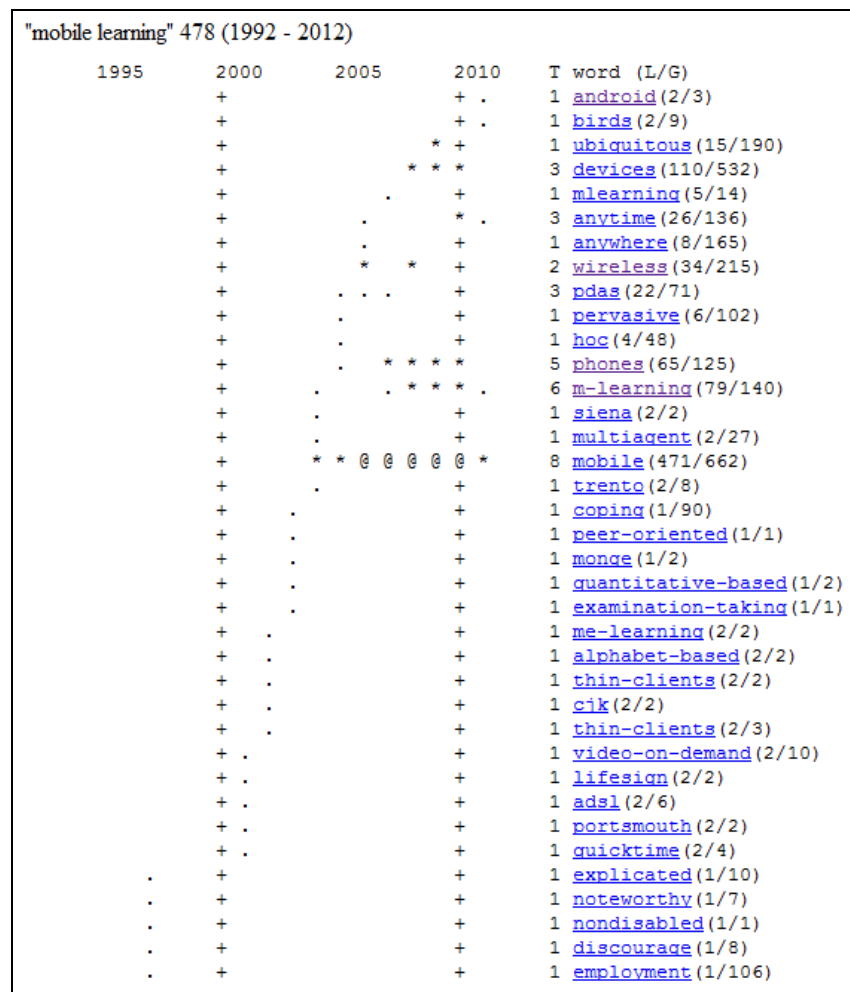


Figure 4. The Milky Way of "mobile learning" feature words

### Comparison of research trends by country

Using the MWRT search engine to compare the research trends of countries initially returned the results as displayed in Figure 5, with the United Kingdom, Spain, and more recently China and Taiwan as countries with strong research. Further detail was obtained using multiple results from the MWRT search engine to build a view of the trend in the number of published articles relating to mobile learning published by country as displayed in Figure 6. The four top ranking countries in mobile learning research by number of articles published are Taiwan, China, the United Kingdom, and Spain. The trends in the number of articles from these countries seem to share a common feature: the number of published articles increases noticeably after the year 2005 when compared to other countries. A novice research student might pose the question: what was the catalyst for the increase in research and published articles? On further investigation it was determined that government public policy might be a significant contributing factor.

In the case of Taiwan, a government driven program promoting e-learning could be linked to the dramatic rise in the number of published articles as described by Hwang (Hwang & Tsai, 2011). In 2005, China's Ministry of Education implemented policies regarding the future focuses of e-learning, thus possibly explaining the sudden increase in research and published articles originating from China (Gilsun, 2006). In the United Kingdom, the Department for Education and Skills' 2005 strategy, "Harnessing Technology", called for the research and development of e-

learning at all levels of education, which would explain the rapid increase in the number of published articles (Department of Education and Skills, 2005). The Spanish Government's Plan Avanza policy, which was ratified in 2005, outlined the advancement of ICT on both a societal and economic level, and is highly likely to be a key force behind the rise in research and in the number of articles published (van Ark, 2011).

"mobile learning" 477 (2000 - 2012)

Top 3 countries of each year with respect to the number of articles that contain the word

T : the number of years when the word was in the top rank  
 L : the number of articles that contain the word in those years  
 G : the total number of articles that contain the word

2000				2005			2010			T word (L/G)			
+	.	.	.	.	.	.	6	13	10	13	7	.	5 c:China(49/1613)
+	.	.	.	.	4	.	9	14	.	12	6	.	5 c:Taiwan(45/856)
+	.	.	.	.	4	5	.	.	.	+	.	.	2 c:Germany(9/712)
+	.	.	.	2	.	.	.	.	7	11	4	.	4 c:Spain(24/819)
+	.	.	.	5	3	5	.	.	.	+	.	.	3 c:Italy(13/511)
+	1	.	.	2	.	5	8	6	10	+	.	.	6 c:United_Kingdom(32/1079)

Figure 5. Number of articles published between 2000 and 2012 by country result

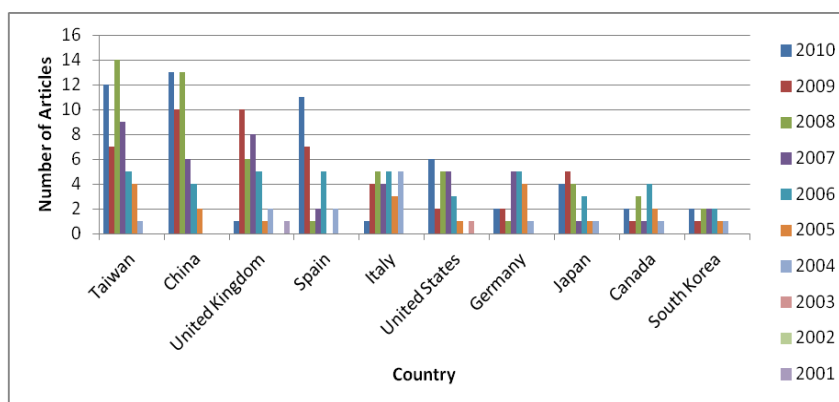


Figure 6. Number of articles published between 2001 and 2010 by country

Using multiple search results from the MWRT search engine to create an international view of research, as demonstrated, may help novice research students identify public or private sector influences that shape research trends within the field of mobile learning.

### Detailed analysis of research trends in China, Japan and Taiwan

(a) China

The results from the analysis of keywords from China are displayed in Figure 7. A total of 55 articles were returned that included possible trend keywords. Of particular interest are keywords which relate to the way learning is undertaken. Keywords, such as: “anytime”, “peer-to-peer”, “ubiquitous” and “self-promotion” suggest that there might be a research trend related to peer based ubiquitous learning.

(b) Japan

In Figure 8 the results of the analysis of research trend keywords from Japan are displayed. A total of 19 articles were returned that included possible research trend keywords. While the total number is significantly less than those returned for China, the degree of keyword variation is similar. Both countries share in common the keyword

“ubiquitous” which has been highlighted as part of a possible trend. The keywords “pdas” and “phones” suggest that hardware device related research may have been an early trend. More recent occurrences may suggest that the trend has moved towards more software based research in mobile learning, such as “movie”, “photos”, and “browsers”.

"mobile learning c:China" 55 (1992 - 2012)				
1995	2000	2005	2010	T word (L/G)
+			. .	2 <a href="#">anytime</a> (4/136)
+			. + .	2 <a href="#">wap</a> (3/11)
+		. .	+	2 <a href="#">wireless</a> (8/215)
+		. * . *	.	5 <a href="#">mobile</a> (49/662)
+		.	.	2 <a href="#">peer-to-peer</a> (3/78)
+		.	. +	2 <a href="#">ubiquitous</a> (4/190)
+		. .	+	2 <a href="#">manifested</a> (2/43)
+		. .	+	2 <a href="#">self-promotion</a> (2/3)
+		. .	+	2 <a href="#">persistence</a> (2/40)
+		.	. +	2 <a href="#">encapsulated</a> (2/21)
+		. .	+	2 <a href="#">migration</a> (2/37)

Figure 7. Research trends of China

"mobile learning c:Japan" 19 (1992 - 2012)				
1995	2000	2005	2010	T word (L/G)
+		.	.	2 <a href="#">browsers</a> (2/116)
+		.	. +	2 <a href="#">photos</a> (2/14)
+		.	. +	2 <a href="#">grasping</a> (2/42)
+		.	. +	2 <a href="#">movie</a> (2/37)
+		.	. +	2 <a href="#">everyone</a> (2/34)
+		. .	+	2 <a href="#">limited</a> (2/992)
+		. . .	+	3 <a href="#">phones</a> (4/125)
+		. .	+	2 <a href="#">pdas</a> (2/71)
+		. .	+	2 <a href="#">mobile</a> (2/662)
+		. .	+	2 <a href="#">ubiquitous</a> (2/190)

Figure 8. Research trends of Japan

### (c) Spain

The results from the analysis of research trend keywords from Spain are displayed in Figure 9. A total of 32 articles were returned, which, along with the keyword variation, is moderate when compared to the previous results for China and Japan. The resulting keywords may suggest that the trends are mainly based around hardware related research, such as “desktops”, “handhelds”, “devices”, “pdas”, and “phones”.

"mobile learning c:Spain" 32 (1992 - 2012)				
1995	2000	2005	2010	T word (L/G)
+		. .	+	3 <a href="#">phones</a> (3/125)
+		. .	+	2 <a href="#">pdas</a> (2/71)
+		. .	+	3 <a href="#">m-learning</a> (6/140)
+		.	* .	3 <a href="#">mobile</a> (20/662)
+	.	.	+	2 <a href="#">domosim-tpc</a> (2/3)
+	.	.	+	2 <a href="#">pretenders</a> (2/7)
+	.	.	+	2 <a href="#">metaphors</a> (2/60)
+	.	.	+	2 <a href="#">devices</a> (6/532)
+	.	.	+	2 <a href="#">desktops</a> (2/67)
+	.	.	+	2 <a href="#">wireless</a> (3/215)
+	.	.	+	3 <a href="#">convergence</a> (4/124)

Figure 9. Research trends of Spain



#### (d) Taiwan

Analysis results from the search for research trend keywords from Taiwan are displayed in Figure 10. The total number of returned articles is an impressive 58, with a wide range of different possible research trend keywords when compared to the previously reviewed countries, China, Japan, and Spain. Of particular interest is the rapid rise in the number of articles published around 2004 to 2005, as previously discussed, which could possibly be attributed to Taiwan's push into e-learning as described by Hwang (Hwang & Tsai, 2011).

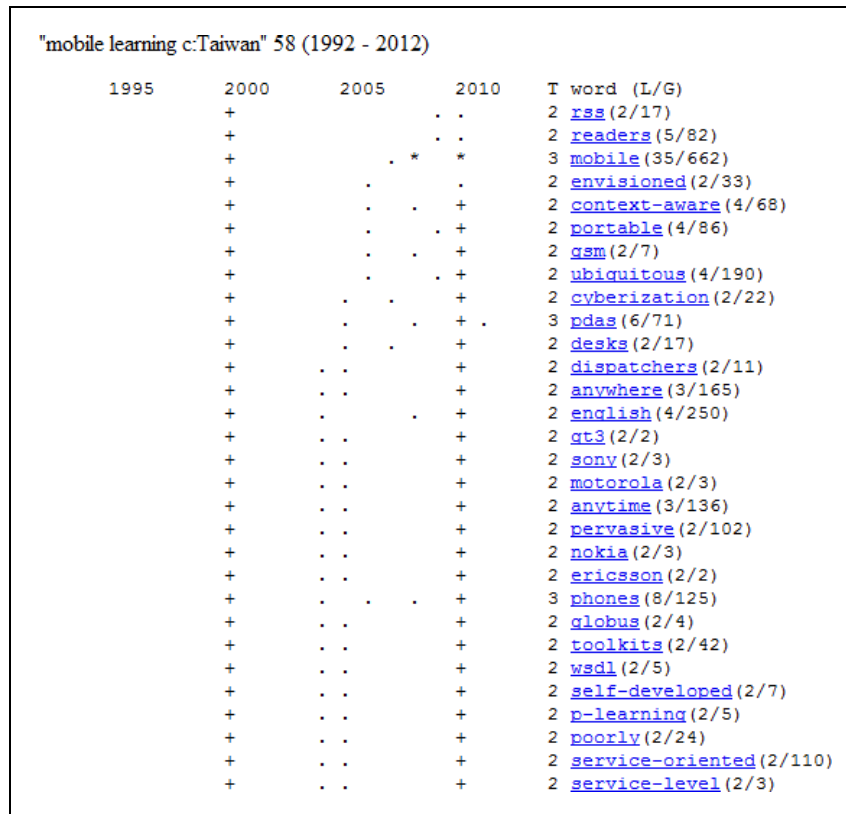


Figure 10. Research trends of Taiwan

## Experiment design

To evaluate the effectiveness of the approach in helping students recognize the progression of trends and keyword transitions for specific learning topics, an experiment has been conducted on a college Information Management course. In the following subsections, the design of the experiment is described in detail.

### Participants

The participants were two classes of sophomore students of a university in northern Taiwan. A total of sixty-nine students participated in the study, including twenty-five females and forty-four males with an average age of 20. One class was assigned to be the experimental group, and the other was the control group. The experimental group included thirty-six students (twelve males and twenty-four females), while the control group had thirty-three students (thirteen males and twenty females). In order to avoid the influence of different instructors on the experimental results, the two classes were taught by the same instructor. The experimental group learned with the MWRT search engine, while those in the control group learned with a search engine with a conventional interface (i.e., displaying the searched results by showing a list of document titles and abstracts). Both groups received the information-searching and summarizing instructions for analyzing the research trends of e-learning articles before the learning activity.



## Measuring tools

The measuring tools of this study include a pre-test, a post-test, and several questionnaires for measuring students' cognitive load, satisfaction and acceptance of using the search engines.

The pre- and post-test were developed by two experienced teachers who had taught the Information Management course for a number of years. The pre-test aimed to ensure that both groups of students had the equivalent basic knowledge of using computer networks and search engines. It consisted of forty multiple-choice items. The post-test aimed to evaluate the students' learning achievement (i.e., their knowledge of the progression of trends and keyword transitions for specific learning topics) after participating in the learning activity. It consisted of twenty multiple-choice items related to the trends of the specific topic investigated during the learning activity. Both the perfect scores of the pre-test and post-test were 100.

The cognitive load questionnaire was modified from the cognitive load measure developed by Sweller, van Merriënboer and Paas (1998). It consisted of 4 items in the two dimensions of "mental load" and "mental efforts" on a six-point Likert scale, where "6" represented "strongly agree" and "1" represented "strongly disagree." The Cronbach's  $\alpha$  value of the cognitive load questionnaire was .79, showing good reliability in internal consistency.

The questionnaire of technology acceptance of using search engines was developed based on the questionnaire developed by Hwang, Wu, Tseng, and Huang (2011). It consisted of 13 items, including 7 items for perceived ease of use and 6 items for perceived usefulness with a seven-point Likert scale. The Cronbach's  $\alpha$  values of the two questionnaire dimensions were .89 and .94, respectively. In addition, three open-ended items were used to collect opinions from the students after experiencing using the search engine.

## Experimental procedure

The experiment was conducted on a unit of an Information Management course in a university. The objective of the unit is to teach students the notation of literature surveys, including the progression of trends and keyword transitions for specific learning topics.

As shown in Figure 1, before the experiment, the students took a pre-test for evaluating their basic knowledge of using computer networks and search engines. Before the learning activity, an orientation was given to introduce the learning tasks and the learning system. Following that, a 150 minute learning activity was conducted. Both the experimental and control groups were asked to search for information to answer a series of questions related to a research issue, such as "What is the trend of research methods and tools related to this issue in the past decade" and "Compare the studies conducted by researchers of five countries and report the differences and similarities between their studies in terms of subjects, methods and tools." The students in the experimental group learned with the MWRT search engine, whereas those in the control group learned with the conventional search engine. After the learning activity, a post-test was conducted; moreover, the students were also asked to complete the cognitive load and the technology acceptance questionnaires on completion of the post-test.

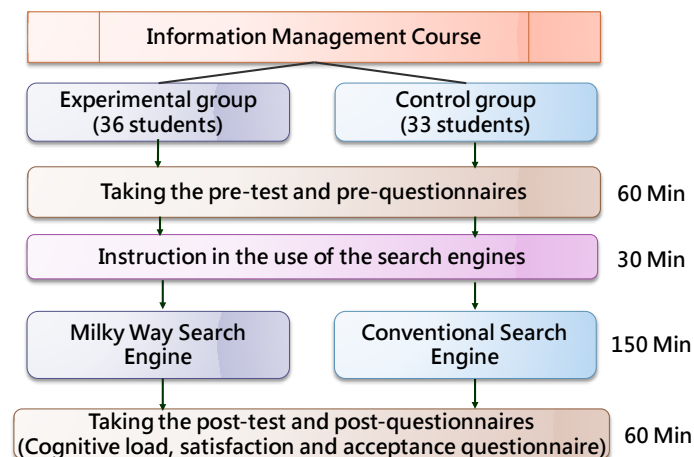


Figure 1. Diagram of experiment design

## Experimental results

In the present study, the collected data were first examined by descriptive statistics to explore the group means, standard deviations and numbers. Then, t-tests and analysis of covariance (ANCOVA) were conducted to examine the effects of using the proposed approach on students' learning achievements in recognizing research trends and their perceptions of the learning activity. In addition, the technology acceptance degrees of the students were analyzed as well.

### Analysis of learning achievement

The aim of this study was to examine the effectiveness of the MWRT search engine in helping students identify possible trends in specific fields. From the pre-test, it was found that the mean values and standard deviations were 62.99 and 7.09 for the experimental group, and 60.08 and 6.97 for the control group. The t-test result ( $t = 1.71$ ,  $p > .05$ ) shows that there was no significant difference between the two groups; consequently, it is concluded that the two groups had equivalent prior knowledge before the learning activity, as shown in Table 1.

Table 1. Descriptive data and t-test result of the post-test results

Variable	Group	N	Mean	S.D.	<i>t</i>
Pre-test	Experimental group	36	62.99	7.09	1.71
	Control group	33	60.08	6.97	

After the learning activity, the analysis of covariance (ANCOVA) was used to test the difference between the two groups by using the pre-test scores as the covariate and the post-test scores as dependent variables. Table 2 shows the ANCOVA results of the post-test. The adjusted mean value and standard error of the post-test scores were 83.97 and 1.52 for the experimental group, and 77.98 and 1.59 for the control group. According to the results ( $F = 7.24$ ,  $p < .01$ ), there was a significant difference between the two groups; that is, the students who learned with the MWRT search engine showed significantly better learning achievements than those who learned with the conventional search engine, implying that the MWRT search engine is helpful to students in recognizing the progression of trends for specific learning topics.

Table 2. Descriptive data and ANCOVA result of the post-test results

Variable	Group	N	Mean	S.D.	Adjusted Mean	<i>F</i>
Post-test	Experimental group	36	84.56	1.52	83.97	7.24**
	Control group	33	77.33	1.59	77.98	

\*\*  $p < .01$

### Analysis of cognitive load

In addition to learning achievement, it is worth investigating whether the new search engine increases the learning pressure of the students. Consequently, the cognitive load measure was used to compare the pressure of the students from two aspects; that is, mental load and mental effort. The former is concerned with pressure caused by the amount of information presented to the learners, while the latter is related to pressure caused by the way of structuring the information (Sweller, van Merriënboer, & Paas, 1998; Hwang & Chang, 2011).

Table 3 shows the t-test results for the mental load and mental effort scores of the two groups. In terms of mental load, the mean and standard deviation were 3.10 and 1.06 for the experimental group, and 2.91 and 0.97 for the control group; in terms of mental effort, the mean and standard deviation were 3.69 and 1.27 for the experimental group, and 3.17 and 1.16 for the control group. The t-test results show that there was no significant difference between the two groups in either dimension, implying that the MWRT search engine did not increase the cognitive load of the students although its user interface does seem complicated.

*Table 3. Descriptive data and t-test result of the cognitive load*

Variable	Group	N	Mean	S.D.	<i>t</i>
Mental Load	Experimental group	36	3.10	1.06	0.76
	Control group	33	2.91	0.97	
Mental Effort	Experimental group	36	3.69	1.27	1.80
	Control group	33	3.17	1.16	

### Analysis of technology acceptance

Table 4 shows descriptive statistics of the feedback from the students in the experimental group and control group to the questionnaire items of technology acceptance of using the search engines. For the aspect of perceived usefulness, the means were 3.80 for the experimental group and 3.11 for the control group; moreover, the t-test result ( $t = 3.04$ ,  $p < .01$ ) for this dimension shows that there was significant difference between the two groups, implying that the students in the experimental group showed strong recognition of the usefulness of the MWRT search engine in helping them conduct trend surveys.

On the other hand, the t-test result ( $t = 1.13$ ,  $p > .05$ ) on the "ease of use" dimension shows that there was no significant difference between the two groups, implying that the MWRT search engine was identified by the students as an easy-to-use system much the same as conventional search engines although it was absolutely new to them.

*Table 4. Results of questionnaire of acceptance and satisfaction with the search engine*

Variable	Group	N	Mean	S.D.	<i>t</i>
Perceived usefulness	Experimental group	36	3.80	0.81	3.04**
	Control group	33	3.11	1.03	
Perceived ease of use	Experimental group	36	3.63	0.81	1.13
	Control group	33	3.39	0.98	

\*\*  $p < .01$

In addition, the students' feedback to the open-ended questions showed that most of the students in the experimental group preferred to use this innovative way to engage in literature retrieval and content analysis in the future. More importantly, some encouraging responses were derived from the students. One of the students stated that "This learning activity is very helpful to novices who have just started to study a topic. It makes my trend survey process more efficient." Another student expressed that, "This way of learning has motivated me to look for more information for understanding the topic in depth." One student further indicated that "In the beginning, I felt it difficult to memorize the functions of this search engine, but soon I learned the skills of using it. I am very happy to have this opportunity to experience using this search engine. It is a useful system and a very innovative way to help us analyze the trends of the topic."

### Conclusions and future work

For the students who are novices in a field, it is essential to conduct literature surveys (Shih et al., 2008). However, as has been indicated by researchers, the existing methods or tools for supporting research trend surveys are inefficient. For example, when doing content analysis, the research topics are first categorized into several tentative categories and sub-categories, and are refined manually and continually by using the constant-comparative method. Moreover, all the selected articles need to be coded manually based on different types of categories referring to their abstract. In addition, highly cited papers need to be further selected to analyze their research participants, research setting, research design and methods (Tsai & Chi, 2011; Tsai & Chiang, 2011). It can be seen that doing such a survey is time-consuming; more importantly, the students spend most of their time categorizing and coding data instead of evaluating the findings and reorganizing their knowledge with conventional approaches.

Jonassen, Carr and Yueh (1998, p. 1) indicated that "Technologies should not support learning by attempting to instruct the learners, but rather should be used as knowledge construction tools that students learn with, not from." They further pointed out that learners need tools that help them to access and process that information; moreover, a

new class of intelligent information search engines can be effective Mindtools to help learners construct knowledge via scanning information resources and organizing the information in a meaningful way.

Conventional search engines, such as Google or Yahoo!, only provide simple information searching and summarizing functions. The learning system proposed in this study not only provide a framework to guide learners determine the keywords and parameters for trend-related information searching, but also presented the searched information in visualized forms (e.g., a time series) via classifying and analyzing the retrieved information, which help learners realize how information can be interpreted and organized for trend analysis. Such a process of searching for, organizing and interpreting information is indeed a form of knowledge construction as indicated by Jonassen and Carr (2000). In the post-test, the students were asked to answer a series of questions related to the trends of a specified research issue that they had not surveyed before; that is, they needed to determine keywords and parameters for collecting information before analyzing the trends of the issue. From the experimental results, it was found that the students in the experimental group showed significantly better performance in answering the trend-related questions of the target issue than those in the control group, implying that MWRT is able to improve their abilities of searching for information as well as interpreting the collected information for trend analysis. From the experimental results, it is concluded that the proposed approach is able to improve the students' awareness of the trends of specific topics. It is also interesting to find that the developed search engine has been identified by the students as an easy-to-use tool much the same as conventional search engines, although it is new to them and its interface seems more complicated. The analysis results of the cognitive load of the students also support the finding that the innovative approach is able to assist students in analyzing the searched data without increasing their mental pressure.

It should be noted that MWRT is not the only system that provides search results by summarizing searched data over time. For example, Google Trends is a well-known system that presents search results as charts in which the horizontal axis represents time and the vertical shows the frequency of people using the keywords in comparison with the total search times. Although both Google Trends and MWRT show the frequency of a term used over time and display the search results as a chart, their functionalities are quite different in several ways. The most crucial differences are the data sources and the mechanism of selecting keywords to be compared. Google Trends analyzes the queries of users and searches for data provided by Internet users; on the other hand, MWRT analyzes scientific articles from a database which can be updated or maintained by researchers or teachers. Moreover, the target keywords for comparison need to be specified by the user in Google Trends, while MWRT extracts the related keywords automatically based on the user's query. Therefore, MWRT is more like a learning tool that allows researchers or teachers to provide data related to a specified learning purpose; in the meantime, it serves as a supporting tool for showing students how to organize the searched data for trend analysis.

In comparison with previous studies that used the content analysis method to identify research trends in the field of e-learning, this "learning by searching" system provides a more efficient and effective environment to foster the research trend survey ability of students. Therefore, such an approach can cope with the problem indicated by researchers that the lack of proper supports or tools might cause students to pay too much attention to the data collection and pre-processing stages, while the evaluation and analysis of the collected data are often ignored or briefly done (Chiou et al., 2009). It also conforms to the conception of Mindtools proposed by Jonassen and Carr (2000) that technology can be used as a Mindtool to engage students in higher order thinking that is necessary for meaningful learning.

On the other hand, the students made some negative comments about the use of MWRT; in particular, they felt that its interface is complex and the symbolic icons are difficult to follow, especially at the beginning of the learning activity. Consequently, in the near future, we plan to improve our system by providing a more easy-to-use interface. Moreover, we also plan to investigate the students' proactive attitude (Schmitz & Schwarzer, 1999), which is a belief in the potential of changes that can be made to improve oneself and one's environment by considering various facets such as resourcefulness, responsibility, values, and vision. Such a personality characteristic is highly related to trend analysis as well as the implications for motivation and action.

## Acknowledgements

This study is supported in part by the National Science Council of Taiwan under contract number NSC 101-2511-S-011-005-MY3.

## References

- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruickshank, K. A., Mayer, R. E., Pintrich, P. R.,...Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives* (Abridged ed.). New York, NY: Longman.
- Bilal, D. (2000). Children's use of the Yahoo! web search engine: I. cognitive, physical, and affective behaviors on fact-based search tasks. *Journal of the American Society for Information Science*, 51(7), 646-665.
- Bloom, B. S. (Ed.), Engelhart, M.D., Furst, E.J., Hill, W.H., & Krathwohl, D.R. (1956). *Taxonomy of educational objectives: Handbook I: Cognitive domain*. New York, NY: David McKay.
- Bruner, J. S. (1967). *On knowing: Essays for the left hand*. Cambridge, Mass: Harvard University Press.
- Chiou, C. K., Hwang, G. J., & Tseng, Judy C. R. (2009). An auto-scoring mechanism for evaluating problem-solving ability in a web-based learning environment. *Computers & Education*, 53(2), 261-272.
- Department of Education and Skills (2005). *Harnessing Technology: Transforming Learning and Children's Services* (DFES-1437-2005). Retrieved from Department of Education Browse publications website: <https://www.education.gov.uk/publications/eOrderingDownload/1437-2005PDF-EN-01.pdf>
- Gilsun S. (2006, April). *eLearning in China – Government Policies and Pilot Universities*. Paper presented at the 10th IACEE World Conference on Continuing Engineering Education (WCCEE), Vienna, Austria.
- Ha, L., De, J., Holden, H., Rada, R. (2009). Literature trends for mobile learning: word frequencies and concept maps. *International Journal of Mobile Learning and Organisation*, 3(3), 275-288.
- Hwang, G. J., & Chang, H. F. (2011). A formative assessment-based mobile learning approach to improving the learning attitudes and achievements of students. *Computers & Education*, 56(4), 1023-1031.
- Hwang, G. J., & Kuo, F. R. (2011). An information-summarising instruction strategy for improving web-based problem solving abilities of students. *Australasian Journal of Educational Technology*, 27(2), 290-306.
- Hwang, G. J., & Wu, P. H. (2012). Advancements and trends in digital game-based learning research: A review of publications in selected journals from 2001 to 2010. *British Journal of Educational Technology*, 43(1), E6-10.
- Hwang, G. J., Chu, H. C., Lin, Y. S., & Tsai, C. C. (2011). A knowledge acquisition approach to developing Mindtools for organizing and sharing differentiating knowledge in a ubiquitous learning environment. *Computers & Education*, 57(1), 1368-1377.
- Hwang, G. J., Tsai, P. S., Tsai, C. C., & Tseng, Judy C. R. (2008). A novel approach for assisting teachers in analyzing student web-searching behaviors. *Computers & Education*, 51(2), 926-938.
- Hwang, G., Tsai, C. C. (2011). Research trends in mobile and ubiquitous learning: a review of publications in selected journals from 2001 to 2010. *British Journal of Educational Technology*, 42(4), E65-E70.
- Hwang, G. J., Wu, C. H., Tseng, Judy C. R., & Huang, I. W. (2011). Development of a ubiquitous learning platform based on a real-time help-seeking mechanism. *British Journal of Educational Technology*, 42(6), 992-1002.
- Jonassen, D. H., & Carr, C. S. (2000). *Mindtools: affording multiple knowledge representations for learning*. In S. P. Lajoie (Ed.), *Computers as cognitive tools. No more walls* (Vol. 2, pp. 165-196). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Jonassen, D.H., Carr, C., & Yueh, H.P. (1998). Computers as Mindtools for engaging learners in critical thinking. *Tech Trends*, 43(2), 24-32.
- Karatas, S. (2008). Interaction in the Internet-based distance learning researches: results of a trend analysis. *The Turkish Online Journal of Educational Technology*, 7(2), article 2.
- Kuo, F. R., Hwang, G. J., & Lee, C. C. (2012). A hybrid approach to promoting students' web-based problem solving competence and learning attitude. *Computers & Education*, 58(1), 351-364.
- Laouris, Y., & Etekleous, N. (2005, October). *We need an educationally relevant definition of mobile learning*. Paper presented at the 4th World Conference on Mobile Learning, Cape Town, South Africa.

- Liu, H. (2008). Learning by Searching. In K. McFerrin et al. (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2008* (pp. 3843-3844). Chesapeake, VA: AACE.
- Schmitz, G. S., & Schwarzer, R. (1999). Teachers' Proactive Attitude: Construct description and psychometric analyses. *Zeitschrift für Empirische Pädagogik*, 13(1), 3-27.
- Shih, M., Feng, J., & Tsai, C. C. (2008). Research and trends in the field of e-learning from 2001 to 2005: A content analysis of cognitive studies in selected journals. *Computers & Education*, 51(2), 955-967.
- Siemens, G. (2004, December 12). *Connectivism: A learning theory for the digital age*. Retrieved from <http://www.elearnspace.org/Articles/connectivism.htm>
- State, P. (2009, November 19). Search engines are source of learning. *ScienceDaily News*. Retrieved from <http://www.sciencedaily.com/releases/2009/11/091119111417.htm>
- Sweller, J., van Merriënboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10, 251-296.
- The Nielsen Company (2011, January 3). *Apple leads smartphone race, while Android attracts most recent customers*. Retrieved from [http://blog.nielsen.com/nielsenwire/online\\_mobile/apple-leads-smartphone-race-while-android-attracts-most-recent-customers/](http://blog.nielsen.com/nielsenwire/online_mobile/apple-leads-smartphone-race-while-android-attracts-most-recent-customers/)
- Tsai, H. H., & Chi, Y. P. (2011). Trend analysis of supply chain management by bibliometric methodology. *International Journal of Digital Content Technology and its Applications*, 5(1), 285-295.
- Tsai, H. H., & Chiang, J. K. (2011). E-commerce research trend forecasting: A study of bibliometric methodology. *International Journal of Digital Content Technology and its Applications*, 5(1), 101-111.
- Tsai, P. S., Tsai, C. C., & Hwang, G. J. (2011). The correlates of Taiwan teachers' epistemological beliefs concerning Internet environments, online search strategies, and search outcomes. *The Internet and Higher Education*, 14(1), 54-63.
- Tseng, Judy C. R., & Hwang, G. J. (2007). Development of an automatic customer service system on the internet. *Electronic Commerce Research and Applications*, 6(1), 19-28.
- Tseng, Judy C. R., Hwang, G. J., Tsai, P. S., & Tsai, C. C. (2009). Meta-analyzer: A web-based learning environment for analyzing student information searching behaviors. *International Journal of Innovative Computing, Information and Control*, 5(3), 567-579.
- van Ark, B. (2011, July). *The linked world: How ICT is transforming societies, cultures and economies*. Retrieved from the Conference Board website: <http://www.conference-board.org/subsites/index.cfm?id=9077>

## Scaffolding Environment for e-Learning through Cloud Computing

**Marijana Despotović-Zrakić<sup>1\*</sup>, Konstantin Simić<sup>2</sup>, Aleksandra Labus<sup>3</sup>, Aleksandar Milić<sup>4</sup>  
and Branislav Jovanić<sup>5</sup>**

<sup>1</sup>University of Belgrade, Faculty of Organizational Sciences, 11000 Belgrade, Jove Ilića 154, Serbia // <sup>2</sup>University of Belgrade, Faculty of Organizational Sciences, 11000 Belgrade, Jove Ilića 154, Serbia // <sup>3</sup>University of Belgrade, Faculty of Organizational Sciences, 11000 Belgrade, Jove Ilića 154, Serbia // <sup>4</sup>University of Belgrade, Faculty of Organizational Sciences, 11000 Belgrade, Jove Ilića 154, Serbia // <sup>5</sup>University of Belgrade, Institute of Physics, 11000 Belgrade, Pregrevica 118, Serbia // maja@elab.rs // kostja@elab.rs // aleksandra@elab.rs // milic@elab.rs // brana@ipb.ac.rs

\*Corresponding author

(Submitted February 01, 2012; Revised July 06, 2012; Accepted October 04, 2012)

### ABSTRACT

This paper presents a new approach to the integration of web services for e-learning through cloud computing infrastructure. The primary goal is to improve the performance of the existing e-learning system where group adaptation is applied. The proposed method includes an integration of information, processes, applications and human resources in the e-learning system. Integration is based on a web application we developed for managing the learning environment. The application enables an integration of the Moodle, the OpenLDAP directory for managing user accounts, and the OpenNebula toolkit for managing a cloud computing infrastructure. The major advantage of the suggested approach is that the existing system for adaptive e-learning does not need any changes, while it becomes more available and suitable for ubiquitous learning. We evaluated the implemented environment in an undergraduate course at the E-Business Lab, University of Belgrade.

### Keywords

E-learning infrastructure, Cloud computing, Environment for adaptive learning

### Introduction

In recent years e-learning has grown into a widely accepted way of learning, and the usage of the global network is inevitable in every education process. Ubiquitous learning integrates wireless, mobile and context awareness technologies in order to detect the situation of the learners and provide more seamless adaptive support beyond formal learning process (Shih, Chu, Hwang, & Kinshuk, 2011; Hwang, Chih-Hsiang, Tseng, & Huang, 2011, El-Bakry & Mastorakis, 2009; Yang, 2006). In order to support modern pedagogical approaches, as well as a variety of heterogenic learning resources within courses, ubiquitous learning environments need to be based on a powerful IT infrastructure. At the same time, in order to be efficient, ubiquitous learning environments need to be based on learning management systems (hereinafter: LMS) and integrated into an existing e-learning environment of educational institutions.

LMSs are powerful integrated systems that support a number of activities performed by teachers and students during the e-learning process (Hauger & Kock, 2007; Kahiigi, Ekenberg, Hansson, Tusubira, & Danielson, 2007). In most cases, LMS users belong to heterogeneous groups with different, sometimes even adverse, individual characteristics and needs. The adaptation of e-education systems to an individual or to a group based on their characteristics, expectations, knowledge, and preferences is nowadays inevitable (Paramythis & Loidl-Reisinger, 2004). Since systems for adaptive e-learning are becoming more complex, educational institutions need new solutions for deploying scalable and reliable environments for adaptive e-learning (Aroyo, Dolog, Houben, Kravcik, Naeve, Nilsson, & Wild, 2006).

In this paper we set out to explore how an environment for adaptive e-learning can be provided through cloud computing. The main goal of the research is to provide a highly effective, scalable cloud computing services for end-users. The research context of this paper is focused on the e-learning processes within the E-Business Lab, the University of Belgrade.

## Related work

Cloud computing (hereinafter: CC) is an abstract, scalable and controlled computer infrastructure that hosts applications for the end-users. CC is an area of computing that refers to providing customers with highly scalable IT capacities as a service via the Internet (Sultan, 2010). Services and data coexist in a shared and dynamically scaled set of resources (Srinivasa, Nageswara, & Kumari, 2009). Virtualization is one of prerequisites for the realization of CC (Dong, Zheng, Yang, Li, & Qiao, 2009). It allows for an efficient usage of resources, because several virtual machines (hereinafter: VM) can operate on one physical machine (Jin, Liao, Wu, Shao, & Luo, 2008). CC is an infrastructure that can bring a new value to an e-learning system, as educational services can be delivered in a reliable and efficient way. It also provides a suitable environment for ubiquitous learning activities. As a result, efforts to introduce CC in e-learning environment have been initiated over the last couple of years and are ongoing across the world. However, shifting from a traditional IT infrastructure to a cloud based infrastructure is a complex task for an educational institution (Reich, Hubner, & Kuijs, 2012).

The background for this research can be found in the works that explain the benefits of using CC infrastructure in e-learning (Dong, Zheng, Yang, Li, & Qiao, 2009; Zhang, 2010). In (Pocatilu, Alecu, & Vettrici, 2009) the authors explain the key benefits of the CC application in e-learning: improved improbability, virtualization, centralized data storage, costs. Doelitzscher et al. (2011) notice that universities' information systems have periods of intensive usage, when available hardware and software resources have to be efficiently used. Dong et al. (2009b) give a good theoretical and practical basis about using CC infrastructure in e-learning. They presented main layers of an e-learning ecosystem: infrastructure, content and application. Further, they propose four ad hoc modules for managing the e-learning ecosystem: monitoring, module, policy module, arbitration module, and provision module. The authors emphasize the importance of cloud services integration into an e-learning system. The architecture of BlueSky cloud framework includes six layers: user interface, application, common service, capability, data information, virtual infrastructure (Dong, Zheng, Qiao, & Shu, 2009). However, most of the research in the area of CC application in e-learning systems did not involve any learning domain; instead, the research mainly focused on the investigation of motivation, advantages, and technical possibilities, but no concrete implementation of the proposed models could be found.

Considering the existing IT infrastructure in an educational institution, the CC paradigm can be implemented with various approaches. An important issue is to choose the deployment model appropriate for the educational institution. Depending on the type of ownership of physical resources and infrastructure the following deployment models of cloud computing can be developed: a private cloud, a public cloud, a hybrid cloud, and a community cloud. The characteristics of each deployment model from the standpoint of infrastructure are management, ownership and location; from the point of users these are access rights to cloud resources (Mell & Grance, 2011). Researches and case studies pointed out the most common approaches, not only within universities, but in the other fields of CC solutions, are private and public cloud (Jin, Ibrahim, Bell, Gao, Huang, & Wu, 2010). Accordingly, these two models are discussed and analyzed in the following text.

Public clouds are owned and operated by third parties, i.e. cloud service providers. Main advantage of a public cloud is that they may be larger than an enterprise cloud, thus providing the ability to scale seamlessly, on demand. Educational institutions do not need to invest and house large IT infrastructures for educational and research purposes. The first concrete implementation of the CC concept in e-learning included hosting on providers' infrastructure. In (Malan, 2010) the author presented an example of a computer science course realized within the public cloud. Using the Amazon Elastic Compute Cloud (EC2), a load-balanced cluster of VMs for 330 students was developed at Harvard College. Other universities provided a variety of cloud based e-learning services via well known cloud providers such as: Google, Microsoft, Amazon, etc. For instance, the University of Westminster used the Google Apps for Education platform that provides a set of Google services (e-mail, document management and collaboration tools) for the realization of a teaching and learning process (Sultan, 2010). A similar approach was presented in (Herrick, 2009) where Colorado State University migrated to Google Apps for Education. However, in these solutions, there are no tools and interfaces for teachers to supervise students; no integration with existing IT infrastructures has been performed, and no adaptive learning services have been developed.

Opposite the public cloud model, a private cloud model enables educational institutions to have complete control of services, data security, applications and resources that are provided to their users. Recently, the number of private cloud based solutions within e-learning systems has been significantly increased (Liang & Yang, 2011). Depending



on how technology is provided and used, these solutions implement one or more CC service models (Costanzo, Assuncao, & Buyya, 2009): infrastructure as a service (IaaS), platform as a service (PaaS) and software as a service (SaaS). IaaS aims to provide computing resources or storage as a service to users. In this model users install operating system on the machines as well as their application software by themselves. PaaS model enables programming language execution environment for users. This model allows users to develop and deploy their own software solutions. SaaS model provides specific software that runs on a cloud infrastructure. The users of these services do not control or manage underlying infrastructure and application platform. Table 1. provides an overview of existing implementations of private cloud based e-learning solutions considering cloud service models.

*Table 1. Overview of research on e-learning solutions based on private cloud*

Paper	IaaS	PaaS	SaaS	LMS	Educational services	Integration
Wenhong, Sheng, & Guoming (2010)		✓				
Dong, Zheng, Qiao, & Shu (2009)	✓					
Doelitzscher, Sulistio, Reich, Kuijs, & Wolf (2011)	✓	✓	✓	✓	✓	✓
Vouk et al. (2008)		✓	✓		✓	
Costanzo, Assuncao, & Buyya (2009)	✓					
Rajam, Cortez, Vazhenin, & Bhalla (2010)			✓		✓	✓
Liang & Yang (2011)	✓		✓		✓	

IaaS solutions are presented in (Costanzo, Assuncao, & Buyya, 2009; Dong, Zheng, Qiao & Shu, 2009). Costanzo et al. (2009) describe an environment for running applications on top of interconnected infrastructures called InterGrid. The experiments showed significant improvement of the overall system performances. BlueSky cloud framework (Dong, Zheng, Qiao & Shu, 2009), developed by Xi'an Jiaotong University in China, allows for physical machines to be virtualized and allocated on-demand for e-Learning systems.

Wenhong et al. (2010) proposed a framework for managing PaaS in a virtual Cloud computing lab that implements the user management, resource management and access management. Vouk et al. (2008) describe the Virtual Computing Laboratory, a CC solution that was developed within the North Carolina State University, USA. It enables students to reserve and access VMs with basic images or specific applications environments. This platform has been used by a huge number of users.

Delivering particular software applications and services for users of e-learning ecosystems (SaaS) is the most frequent CC service model. Various types of applications such as word processing, presentations, spreadsheets, databases and more can all be accessed through a web browser, while the content is hosted in the cloud (Liang & Yang, 2011). Virtual Computing Laboratory (Vouk et al., 2008) provides students with variety of specific applications (Matlab, Solidworks, Maple, etc.) for different area of study. In (Rajam, Cortez, Vazhenin & Bhalla, 2010) the authors proposed e-learning computational cloud (eLC<sup>2</sup>) that enables e-learning task management. Main goal of the eLC<sup>2</sup> solution was to enhance collaboration among the main actors of learning activities. Virtual Personalized Learning Environment (VPLE) described in (Liang & Yang, 2011) integrates multiple disparate application systems of schools into one integrated online system. The learning content providers can dynamically register new applications at any time and students can personalize the contents of their learning environments.

In (Doelitzscher, Sulistio, Reich, Kuijs, & Wolf, 2011) the authors introduce a comprehensive private cloud solution CloudIA infrastructure that provide IaaS, SaaS, PaaS with respect to requirements and needs of e-Learning and collaboration. CloudIA was developed within Hochschule Furtwangen University. The CloudIA IaaS system enables the creation of a VM by choosing a base image and post-installing software packages selected by the user on the fly. PaaS approach allows students to use VMs with predefined packages and software for their courses. Collaboration Software (CollabSoft) has been adapted to a SaaS model.

Although solutions found in the literature add new value to e-learning systems and enable numerous services and features, the main challenge of using CC in e-education is integration into existing infrastructures of universities. The framework proposed in (Wenhong, Sheng, & Guoming, 2010) provides a comprehensive set of features for managing a virtual computing lab, but it is rather general and not adjusted to the needs of an e-learning system. Also, the approach suggested in (Vouk et al., 2008), where students are provided with predefined VMs, does not take into account any learning environment. Namely, key component of a modern e-learning system is a software solution for

learning management. Most of the universities organize e-learning courses using an LMS. Accordingly, if CC services are not integrated with LMS, they will not add significant improvement to an e-learning system.

In works (Liang & Yang, 2011; Rajam, Cortez, Vazhenin & Bhalla, 2010; Doelitzscher, Sulistio, Reich, Kuijs, & Wolf, 2011; Vouk et al., 2008), some educational services are delivered using CC: collaboration, communication, learning content management. However, only few partial implementations of CC services integration with LMS can be found in literature. For instance, CloudIA provides an on-demand Online Learning And Training LMS as a service for collaboration (Doelitzscher, Sulistio, Reich, Kuijs, & Wolf, 2011). Further, eLC<sup>2</sup> architecture includes different components, such as project management, external task processing, integrated through web services.

Given the existing body of knowledge, the key reason for implementing a private cloud in the E-business Lab at University of Belgrade is that the existing IT infrastructure could be used in a more effective way. Private cloud solution would enhance realization of research and educational processes and enable seamless integration of the CC paradigm in the e-learning system. Also, the E-business Lab already employs experienced technical and teaching staff capable of developing, managing and exploiting the CC infrastructure. Since our courses are related to computer science, the main requirement for teaching and learning process realization was to provide students with an appropriate environment for application development (PaaS model), as well as with sophisticated software tools (SaaS model). Accordingly, the focus of our research is to enable plethora of services and applications for e-learning courses, but not hardware resources (IaaS). Thus, the main idea of our research was to provide SaaS and PaaS on private cloud, and to seamlessly integrate them in Moodle LMS.

## **Integration of services for e-learning through cloud computing**

### **Adaptive e-learning in the e-business lab**

The essential component of an adaptive e-learning infrastructure in the e-Business Lab is the Moodle LMS. The Moodle is flexible when it comes to extending with new components and integration with other systems and technologies necessary for adaptation. The system for e-learning within the department provides students with courses from different areas and aims to support both formal and informal learning. Through the Moodle, the system enables a realization of common formal learning activities and tasks (assignments, tests, learning resources, etc.). At the same time, in order to support ubiquitous learning, we have developed various applications and services (mobile applications, edutainment activities, social media activities, etc.) described in Barać et al. (2011) and Vulić et al. (2012).

The adaptivity in the e-learning system is realized using the model described in (Despotovic-Zrakić, Marković, Bogdanović, Barać, & Krčo, 2012). Group personalization of courses has been implemented with respect to Felder-Silverman learning styles. Different learning materials and activities are created for each group. All student activities are tracked using the Moodle LMS. Personalization is reflected in adapted resources, activities, tasks and tools for communication. The realization of group personalization was effective, but new problems arose. Each adaptive course contains many different resources, tools and applications that require adequate hardware and software resources. If these resources could be available on demand, a student would need only an Internet connection and a web browser. Therefore, we need to develop an environment for adaptive e-learning where some resources would be provided through personalized Moodle courses, while others would be provided through the CC infrastructure on demand. The CC infrastructure should enable the hosting of all resources and services with the best possible performance (Vujin, 2011). This approach enables the integration and synchronization of all components in the scope of a learning process into a single system. In the text that follows, concrete problems and requirements of each of our courses are described shortly.

The basic idea of the *Internet technologies* course is to learn the key concepts in the development of web applications. Students have to create advanced web applications enriched with JavaScript, JQuery, XML technologies, as well as web services. In order to implement their projects, it is necessary to provide different operating systems (Linux, Windows), development platforms (NetBeans, Eclipse, Dreamweaver, Visual Studio), software tools, database management tools (SQL server, MySQL), web servers (Apache, IIS), all modern browsers, etc. Therefore, various combinations of these technologies and platforms should be available to students on their VMs.

The *E-business* course deals with key concepts in design and implementation of e-business projects. In the scope of their activities within the course, students create web presentations and e-commerce web sites, use ERP and CRM solutions, tools for web design, etc. All these activities require appropriate platforms to be installed, as well as web hosting services. Students should be provided with a VM created in accordance with their needs.

In the scope of the subject *M-business*, modern trends and principles in the area of m-business applications, as well as the Internet of things, are studied. Students develop applications for various mobile platforms. Different platforms and tools that require high-performance hardware are necessary (Android SDK tools, Eclipse, NetBeans). Various technologies are used, such as: RFID, GPS, wireless networks, etc.

The main goal of the *Simulation and simulation languages* course is to teach students the basic and advanced concepts of computer simulation. The course includes three areas: continuous simulation, discrete simulation and 3D modeling. According to this, students need different software solutions for learning simulation.

### Integration model

The development of a typical system for e-learning includes: the implementation of LMS, the integration of Internet services in a network of educational institutions and a business information system. The integration of components of the system is realized using multiple layers:

- Human resource integration – students, teachers and other participants in the learning process can access the system and can communicate from any location.
- Information integration – the system enables gathering heterogeneous, unstructured data, while users can access structured data.
- Process integration – adaptive e-learning processes are integrated using web services.
- Application integration – the integration is realized at the application level on cloud computing infrastructure.

The method of integration of e-learning services with the cloud computing infrastructure is shown in the Figure 1.

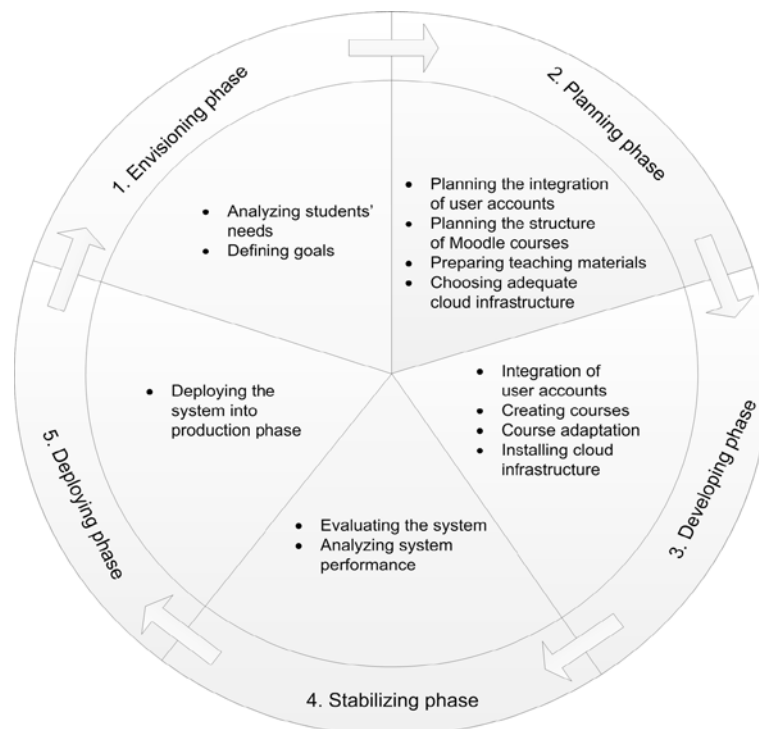


Figure 1. A method for integration of e-learning services with cloud computing

There are five phases in the proposed model. In the first phase, the user accounts are created. The user accounts are stored on LDAP server. The LDAP server is integrated with the user directory of the educational institution where the student accounts are located. In the second phase, the courses are created in the Moodle LMS. Teaching materials are prepared, the activities and the assignments are defined. The necessary software tools for teaching process realization are chosen. The course adaptation process is performed. In the third phase, the VMs with necessary operating systems and software are prepared. Each VM is adapted to students' learning styles and needs at a specific course. Afterwards, the prepared VMs are stored into the CC infrastructure. In the fourth phase, students use the ELABCloud application for VM reservation and its deployment. The application allows students to reserve any of the provided VMs for the Moodle course to which they are enrolled. Students can perform the reservation using a web application. In the fifth and the final phase, teachers and administrators of the system can view and analyze students' results and the performance of the system.

## Implementation of the proposed method

The implementation of the proposed method includes the design and the implementation of the ELABCloud application for managing the cloud computing infrastructure. This application enables the integration of Moodle, OpenLDAP, business information system of educational institution and the CC infrastructure. The application should be deployed within the private cloud infrastructure of the e-Business Lab shown in Figure 2. The key components of the implemented private cloud include: services for access to a virtual environment, services for resource management, services for user account management, a system for distributed file management, and a virtual infrastructure management system. These components allow for an effective work with a VM. VMs are stored in the image repository and can be moved and run on users' demands. Integration of digital identities within heterogeneous e-learning environment has been performed using the LDAP protocol (Zhang & Chen, 2011).

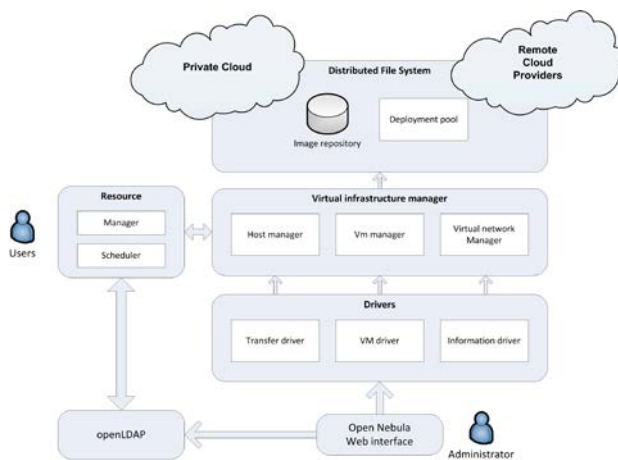


Figure 2. Cloud computing infrastructure in e-Business Lab

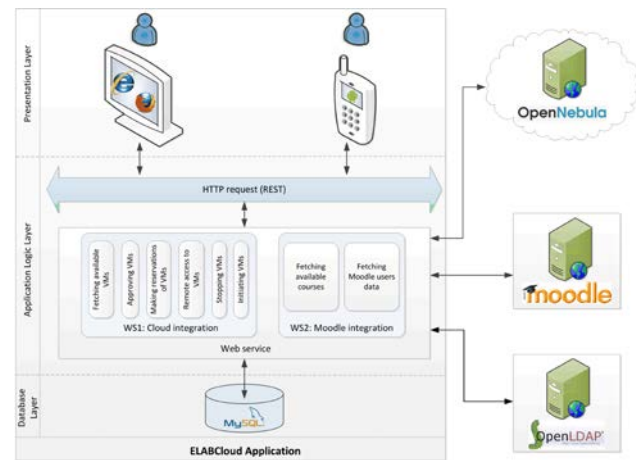


Figure 3. Architecture of the e-learning ecosystem

The ELABCloud application is designed with respect to three-tier architecture (Figure 3). The main components of this complex application include a set of web services and a web application. Via the web application, the user can review, reserve and use VM placed in the cloud infrastructure. The application logic is placed in two web services which integrate the system with the CC infrastructure (WS1) and with the Moodle platform (WS2). User authentication is performed using the LDAP protocol. The same login information is used to log on to the Moodle, the ELABCloud application, the webmail, and other services. The integration with the CC infrastructure is enabled through the following operations: fetching an available VM, approving requests for a VM, making reservations for a VM, remote access to a VM, initiating a scheduled VM and stopping the expired VM.

The process of integration based on the proposed method is shown in the activity diagram in Figure 4.

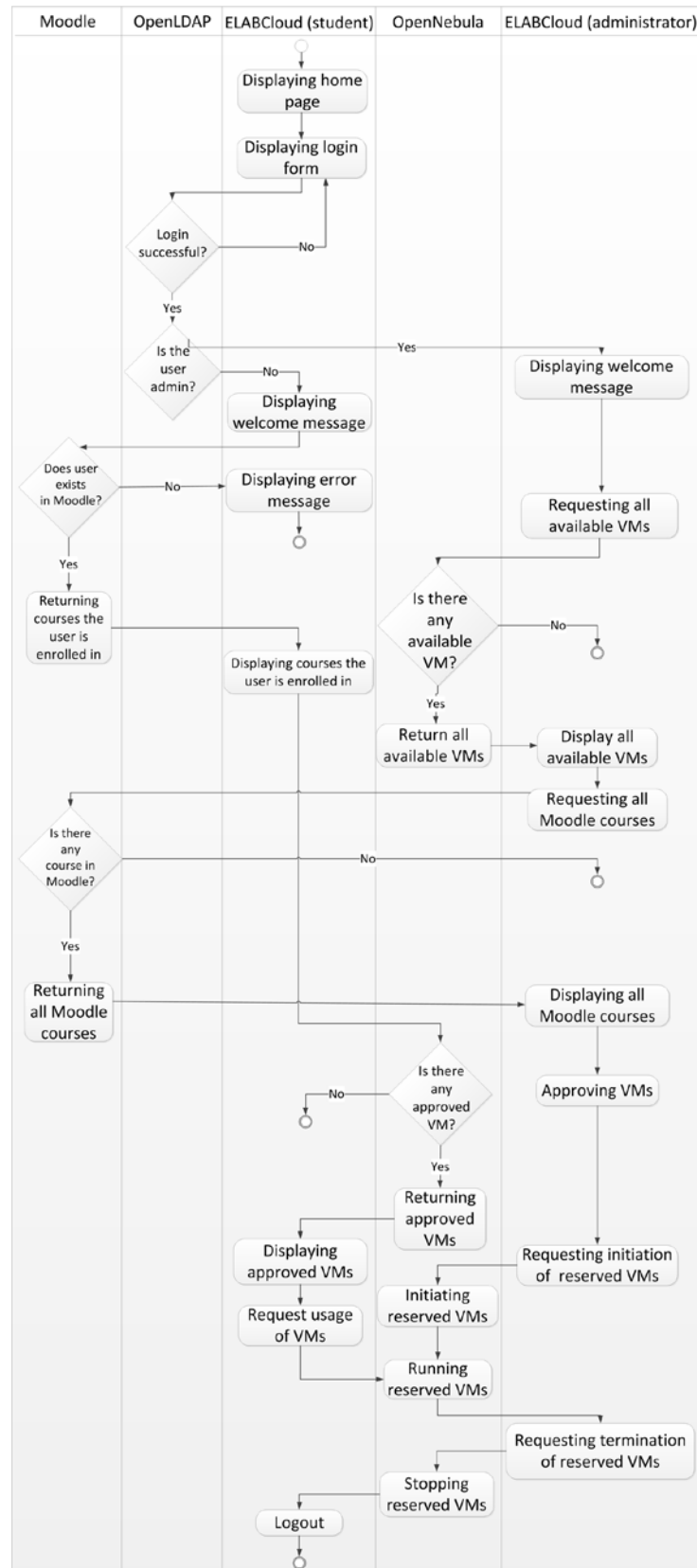


Figure 4. Activity diagram for integration of Moodle and cloud computing infrastructure

## Application of the implemented method

ELABCloud application allows students to reserve any of the enabled VMs for the Moodle course for which they are registered. In this way, VMs will be available to students during the desired period. VMs with all the necessary software have been geared to students' needs depending on the course the student attends. The CC infrastructure with the ELABCloud application enables students to always use the most recent versions of software, regardless of the hardware they own. Students can use this software on any computer with an Internet connection and a web browser. The ELABCloud web application is designed for registered users that have the login credentials stored at the OpenLDAP server.

After a successful login, the student is shown a list of available VMs, depending on the Moodle courses they are enrolled in. The student can make a reservation for a VM, and choose the date and time when the VM will be used (Figure 5). The maximum running period is currently set to 3 hours. The reserved VM can be used in the desired time, from any location and any device with an Internet connection and a web browser. Therefore, this approach provides students with a ubiquitous learning environment.

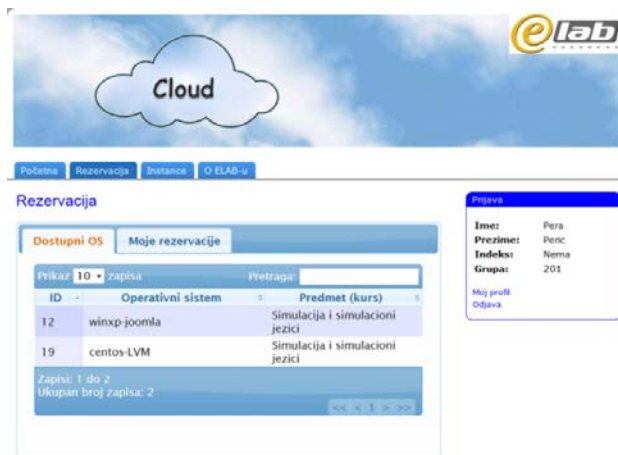


Figure 5. ELABCloud - Virtual machine reservation



Figure 6. ELABCloud – Running virtual machines

When the user logs to the ELABCloud application, they can see and use the started VM (figure 6).

## Evaluation

The evaluation of the described infrastructure for e-learning has been done within the regular teaching activities at the University of Belgrade. The evaluation has been done by comparing the results that students achieved when learning on the CC infrastructure in the school year 2011/12 with those that students achieved when learning on the infrastructure for e-learning that was not cloud based in the school year 2010/11.

In order to better interpret the results, a short qualitative study followed. The intent of the qualitative study was to learn about benefits and problems related to designing and implementing the e-learning infrastructure. The focus was on the e-learning infrastructure necessary for teaching and learning in the area of information and communication technologies, from the teachers' perspective.

Finally, the evaluation of the performance of the developed e-learning infrastructure based on CC has been performed, by using appropriate software tools for performance management.

## Materials and procedure

The evaluation has been done at the Simulation and simulation languages course of study. The course is organized at the fourth year of undergraduate studies in the area of information systems and the Internet. The course is realized

using the combination of traditional classroom-based teaching and e-learning technologies. The Moodle LMS has been used for synchronizing teaching activities and managing online courses. The course deals with the basic concepts and applications of computer simulation and requires the appropriate software tools. Main areas studied in the course are continuous and discrete simulation. For teaching these areas of simulation CSMP/FON and FONWEBGPSS software tools are used respectively, both developed within the E-business Lab (Despotović-Zrakić, Barać, Bogdanović, Jovanić, & Radenković, 2012). CSMP/FON is a desktop application running on a Windows operating system. It is used for creating models of continuous systems and performing simulation. The results are shown in tables and graphs. FONWebGPSS is a web application running on Windows Server 2003, and it requires a set of specific web services for building a discrete simulation model, executing the simulation and visualizing the results of the simulation.

During the school year 2010/11 the CSMP/FON application was provided to students as a desktop application for Windows platform. Each student could download the application from the e-learning course and install it on their personal computer. The application was not available for other operating systems, which arose as a problem for students who did not use Windows. In the school year 2011/12, the CSMP/FON application was provided to students through as a virtual machine through cloud computing. The students could reserve the VM with the CSMP/FON application and access it using a remote connection.

During the school year 2010/11 the FONWEBGPSS application was provided to students as a web application hosted on the Windows server 2003 and the IIS web server. The simulator was integrated in the application through web services, but problems with the speed of simulation were reported. Another problem was related to the scalability and the performance of the application: it could be used by not more than 50 students simultaneously. A higher number of concurrent usages caused serious problems in the performance of the application. In order to overcome these problems in the school year 2011/12, the FONWEBGPSS application was migrated to the newly deployed cloud computing infrastructure. The application was provided to students through the SaaS model. During the semester, students could reserve the VMs with the FONWEBGPSS application in order to practice and prepare for the exam.

While students learned within the cloud infrastructure, the performance of the e-learning system was monitored using the appropriate software tools. Management of different course resources and laboratory data centre, as well as the integration of software solutions and components in the system, were performed through the ELABCloud application.

## Participants

The experiment was conducted on a sample of undergraduate students of the Faculty of Organizational Sciences, the University of Belgrade who attended the Simulation and simulation languages course in school years 2010/11 and 2011/12. The groups' characteristics and treatment are shown in Table 2. The equivalence of the groups was tested using the average grades that students achieved in their previous course of study. The results have shown that there were no significant statistical differences as regards average grades.

*Table 2. Participants and treatment*

School year	N	Learning	Avg.grade	Std.Dev.
2011/12	179	Cloud based	8.11	0.94
2010/11	175	Non-cloud based	8.05	0.99

In the qualitative part of the study, teachers were asked to comment on the developed infrastructure and give their impressions and suggestions. Total of 8 teachers who used the described infrastructure participated.

## Instruments

In order to measure the research results, we have used a knowledge test that students took at the end of semester, and a questionnaire. In the knowledge test, students were required to solve problems in the area of continuous and

discrete system modeling and simulation. The same test has been used to evaluate the students' knowledge in the field of discrete simulation for more than five years.

## Results

A descriptive comparative statistics of results achieved on the knowledge test is presented in Table 3.

*Table 3. Results achieved on knowledge test*

School year	N	Mean	Std.Dev.
2011/12			
CSMP/FON	179	8.07	1.24
FONWEBGPSS	179	8.12	1.31
2010/11			
CSMP/FON	175	7.95	1.32
FONWEBGPSS	175	8.01	1.12

The results show that there are statistically significant differences in results of students in two compared school years for both CSMP/FON and FONWEBGPSS services:

- CSMP:  $F(1,353) = 4.830$  ( $p < 0.05$ )
- GPSS:  $F(1,353) = 4.220$  ( $p < 0.05$ )

Therefore, we can conclude that the application of cloud-based simulation laboratory contributed to the students' results, and that students achieve best results when they learn in the cloud environment.

The results of the qualitative analysis on teachers' opinions on the CC infrastructure can be summarized through the following:

- Most of the teachers reported that in their opinion CC infrastructure had contributed to the higher quality of the e-learning process.
- Teachers responsible for course topics that require more effort for preparing virtual machines stated that although this type of infrastructure brings new value to e-learning, it is time consuming
- Several teachers emphasized that the application of CC infrastructure contributed to higher productivity in realization of e-learning courses. This is because most of the time in practical lessons could be spent on dealing with the main goals of the lesson, while within the previous infrastructure much time was spent on configuring and managing the software environment necessary for learning.
- Teachers stated that the CC infrastructure should be supported by a better reporting system, which would give a better insight on how VMs are used during the learning process.
- Teachers are satisfied with the concept of integration with Moodle, because most of the administration can be done through a single interface.
- Several teachers stated that this type of delivering IT resources is suitable for teaching different courses that include software or information system design and implementation.

Figure 7 shows that the system remained stable, both during the teaching process and the exam taking. The utilization of processing time is less than 60% during the periods of maximum load (at 1:10PM). It is important to notice that the ELABCloud application is the only application used for managing the cloud computing infrastructure. The load of cloud computing infrastructure during the teaching process is given in Figure 7 left. Figure 7 right shows the load of the cloud computing infrastructure during an exam. The diagram shows the user CPU (blue), the system CPU (red) and the idle CPU (orange). The system is not overloaded, even in the period of the highest load (red diagram).



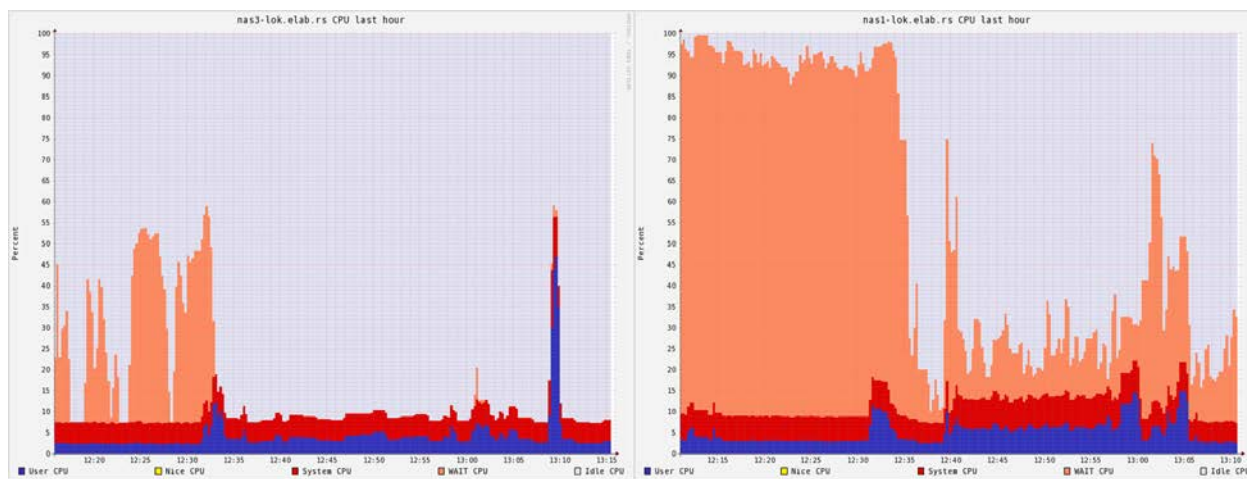


Figure 7. Load of cloud computing infrastructure during the teaching process (left) and Load of cloud computing infrastructure during the exam (right)

## Discussion

Numerous studies have pointed out that CC can provide support and empower e-learning services. The benefits of using CC in e-learning are reflected in: effective usage of resources, improved scalability, integration of different e-learning services, etc. Most of the studies within the area of using cloud computing concept in e-learning deals with theoretical aspects and models (Dong, Zheng, Qiao, & Shu, 2009; Zhang, 2010; Pocatilu, Alecu, & Vetrici, 2009), but only few of them include implementation in real e-learning environment (Doelitzscher, Sulistio, Reich, Kuijs, & Wolf, 2011; Vouk et al., 2008).

This study is oriented towards making an impact on practice. Our e-learning courses cover the area of computer science. Accordingly, our students need variety of operating systems, software tools and applications, development environments. Implemented cloud computing service models (PaaS and SaaS) provide the students with all the necessary software and hardware resources. The approach in this research can be used in any Moodle-centric environment.

Another benefit of applying the described model can be related to cost effectiveness of the private cloud. Although employing a third party cloud computing infrastructure within Harvard College (Malan, 2010) has brought certain benefits such as ability to grow and shrink infrastructure, and an efficient control of learning system performance, this solution might be expensive for many educational institutions. The solution proposed in this paper is related to a more efficient usage of the existing IT resources within educational institutions, and therefore can be economically more acceptable in comparison with using public cloud providers' services.

The significance of the research context is reflected in the integration and synchronization of information, processes, applications and human resources in an adaptive e-learning system. This research demonstrates a measurable improvement of the adaptive e-learning system by developing infrastructure through cloud computing. Adaptive e-learning system described in (Despotovic-Zrakić, Marković, Bogdanović, Barać, & Krčo, 2012) included learning resources and activities adaptation. Integration of CC infrastructure and services enabled delivering complete software environment tailored to students' characteristics, without changes in the system.

As it is depicted in Table 1, majority of the existing solutions for e-learning based on cloud computing, implement some of the cloud service models without any kind of integration with LMS and other learning services. Furthermore, the authors could not find examples of e-learning ecosystems where adaptive learning resources and services were created and delivered through cloud computing. The approach proposed in this paper mainly conforms to the one used in the CloudIA project (Doelitzscher, Sulistio, Reich, Kuijs, & Wolf, 2011), where the e-learning system is based on private cloud that delivers all three cloud service models (IaaS, PaaS, SaaS). Both our and CloudIA solutions include features such as: authentication, resource management, collaboration services, customized on-demand VMs, etc. However, while the CloudIA project mainly focuses on collaboration, the approach proposed in

this paper supports other learning tasks within LMS as well (learning resources and activities, assessment, etc.). Further, VMs are designed with respect to the specific courses requirements and students' needs at the same time.

The results obtained in the experimental part of the paper suggest that students achieve better results when they learn in the cloud environment. Further examinations are necessary in order to determine the specific motives students have for using the cloud infrastructure, as well as the suitability of particular learning resources for delivery through cloud.

Finally, we acknowledge that this approach has some limitations. The main disadvantage of the described approach is that it is based on group personalization and it does not support real time adaptation. Also, this approach is tailored for courses that require extensive usage of diverse hardware and software resources, and might not be suitable for courses with different requirements.

## Conclusion

The main problem treated in this paper is the improvement of the system for e-learning through CC. Our research context is based on improving the system for adaptive e-learning by making it more available without changing the system itself. The described method leads to providing effective and scalable e-learning services, especially suitable for educational institutions that teach computer science or similar subjects. The ELABCloud web application presents a comprehensive solution for an integrated management of CC infrastructure which can be used by both teachers and students. The results of the evaluation suggest that students achieve good results while studying in an e-learning environment improved with CC technologies, which conforms to conclusions reported by (Rajam, Cortez, Vazhenin, & Bhalla, 2010).

With an increasing growth in the number of users, services, educational contents and resources, the approach described in this paper can be used as a model for developing a scalable and reliable infrastructure for e-learning. The model can easily be adjusted and applied in other educational institutions. Although the model is based on the Moodle, the proposed architecture allows for an easy and fast adjustment to other LMSs as well.

Future research will be oriented towards improving the integration with the Moodle, in terms of running a VM directly from a Moodle course. The system for analytics should be improved to provide reports and analyses of the students' results through both Moodle and ELABCloud applications.

## Acknowledgements

The authors of this paper are grateful to MNTRS for financial support grant no. 174031.

## References

- Aroyo, L., Dolog, P., Houben, G-J., Kravcik, M., Naeve, A., Nilsson, M., Wild, F. (2006). Interoperability in personalized adaptive learning. *Journal of Educational Technology & Society*, 9(2), 4-18.
- Barać, D., Bogdanović, Z., Milić, A., Jovanić, B., & Radenković, B. (2011). Developing adaptive e-learning portal in higher education. In J. Martin & C. Baccarani (Eds.), *14<sup>th</sup> Toulon-Verona conference - Organizational excellence in service*, (pp. 128-135). Alicante, Spain: University of Alicante.
- Costanzo, A., Assuncao, M., & Buyya, R. (2009). Harnessing cloud technologies for a virtualized distributed computing infrastructure. *IEEE Internet Computing*, 13(5), 24-33.
- Despotovic-Zrakić, M., Marković, A., Bogdanović, Z., Barać, D., & Krčo, S. (2012). Providing adaptivity in Moodle LMS courses. *Journal of Educational Technology & Society*, 15(1), 326-338.
- Despotović-Zrakić, M., Barać, D., Bogdanović, Z., Jovanić, B., & Radenković, B. (2012). Integration of web based environment for learning discrete simulation in e-learning system. *Simulation Modelling Practice and Theory*, 27, 17-30.

- Doelitzscher, F., Sulistio, A., Reich, C., Kuijs, H., & Wolf, D. (2011). Private cloud for collaboration and e-Learning services: From IaaS to SaaS. *Computing*, 91(1) 23-42
- Dong, B., Zheng, Q., Qiao, M., & Shu, J. (2009a) BlueSky cloud framework: An e-learning framework embracing cloud computing. *Lecture Notes in Computer Science* 5931, 577-582.
- Dong, B., Zheng, Q., Yang, J., Li, H., & Qiao, M. (2009b). An e-learning ecosystem based on cloud computing infrastructure. In A. Ignasio, C. Nian-Shing, Kinshuk, & S. Demetrios (Eds.), *The Ninth IEEE International Conference on Advanced Learning Technologies* 2009 (pp. 125-127). Los Alamitos, CA: IEEE Computer Society.
- El-Bakry, H. M., & Mastorakis, N. (2009). Activation of informal learning with e-learning technology. In S. C. Misra, R. Revetria, L. M. Sztandera, M. Iliescu, A. Zaharim & H. Parsiani (Eds.), *Proceedings of the 8th WSEAS International Conference on Education and Education Technology* (pp. 245-247). Athens, Greece: World Scientific and Engineering Academy and Society.
- Hauger, D., & Kock, M. (2007). State of the art of adaptivity in e-learning platforms. In A. Hinneburg (Ed.), *LWA 2007: Lernen - Wissen - Adaption, Halle, Workshop Proceedings* (pp. 355-360). Halle, Germany: Martin-Luther-University Halle-Wittenberg.
- Herrick, D. (2009). Google this!: using Google apps for collaboration and productivity. *Proceedings of the 37th annual ACM Special Interest Group on University and College Computing Services (SIGUCCS) fall conference* (pp. 55-64). New York, NY: ACM. doi: 10.1145/1629501.1629513
- Hwang, G.-J., Chih-Hsiang, W., Tseng, J. C., & Huang, I. (2011). Development of a ubiquitous learning platform based on a real-time help-seeking mechanism. *British Journal of Educational Technology*, 42(6), 992-1002.
- Jin, H., Ibrahim, S., Bell, T., Gao, W., Huang, D., & Wu, S. (2010). Cloud types and services. In B. Furht & A. Escalante (Eds.), *Handbook of Cloud Computing* (pp. 335-355). New York, NY: Springer.
- Jin, H., Liao, X., Wu, S., Shao, Z., & Luo, Y. (2008). ChinaV: Building Virtualized Computing System. *Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications* (pp. 21-35). Los Alamitos, CA: IEEE Computer Society. doi:10.1109/HPCC.2008.135
- Kahiigi, E., Ekenberg, L., Hansson, H., Tusubira, F., & Danielson, M. (2007). Exploring the e-learning state of art. *Electronic Journal of e-Learning*, 6(2), 77-88.
- Liang, P. H., & Yang, J. M. (2011). Virtual Personalized Learning Environment (VPLE) on the cloud. *Lecture Notes in Computer Science*, 6988, 403-411.
- Malan, D. (2010). Moving CS50 into the cloud. *Journal of Computing Sciences in Colleges*, 25(6), 111-120.
- Mell, P., & Grance, T. (2011). The NIST definition of cloud computing. Retrieved from the Recommendations of the National Institute of Standards and Technology website: <http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf>
- Paramythis, A., & Loidl-Reisinger, S. (2004). Adaptive learning environments and e-learning standards. *Electronic Journal of e-Learning*, 2(1), 181-194.
- Pocatu, P., Alecu, F., & Vetrici, M. (2009). Using cloud computing for E-learning systems. In M. Jha, C. Long, N. Mastorakis, & C. A. Bulucea (Eds.), *Proceedings of the 8th WSEAS international conference on Data networks, communications, computers* (pp. 54-59). Stevens Point, WI: World Scientific and Engineering Academy and Society.
- Rajam, S., Cortez, R., Vazhenin, A., & Bhalla, S. (2010). E-Learning Computational Cloud (eLC<sup>2</sup>): Web services platform to enhance task collaboration. In O. Hoeber, Y. Li & X. J. Huang (Eds.), *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology* (pp. 350-355). Toronto, Canada: Los Alamitos, CA: IEEE Computer Society.
- Reich, C., Hubner, S., & Kuijs, H. (2012). Cloud computing for teaching and learning: Strategies for design and implementation. In L. Chao (Ed.) *Cloud Computing for On-Demand Virtual Desktops and Labs* (pp. 111-125). Hershey, PA: IGI Global.
- Shih, J. L., Chu, H. C., Hwang, G. J., & Kinshuk. (2011). An investigation of attitudes of students and teachers about participating in a context-aware ubiquitous learning activity. *British Journal of Educational Technology*, 42(3), 373-394.
- Srinivasa, R. V., Nageswara, R., & Kumari, K. (2009). Cloud computing: An overview. *Journal of Theoretical and applied Information Technology*, 9(1), 71-76.
- Sultan, N. (2010). Cloud computing for education: A new dawn? *International Journal of Information Management*, 30(2), 101-182.
- Vouk, M., Averitt, S., Bugaev, M., Kurth, A., Peeler, A., Shaffer, H., ... Thompson, J. (2008). "Powered by VCL" - Using Virtual Computing Laboratory (VCL) Technology to Power Cloud Computing. *Proceedings of the 2nd International Conference on*

*Virtual Computing* (pp. 1-10). Retrieved April 11, 2012, from <http://vcl.ncsu.edu/news/papers-publications/powered-vcl-using-virtual-computing-laboratory-vcl>

Vujin, V., (2011). Cloud computing in science and higher education. *Management*, 16(59), 65-70.

Vulić, M., Dadić, J., Radenković, B., Despotović-Zrakić, M. Bogdanović, Z. (2012). Social CRM metrics in e-education. *Metalurgia International*, 17(7), 205-211.

Yang, S. J. H. (2006). Context aware ubiquitous learning environments for peer-to-peer collaborative learning. *Educational Technology & Society*, 9(1), 188-201.

Wenhong, T., Sheng, S., & Guoming, L. (2010) A framework for implementing and managing platform as a service in a virtual cloud computing lab. In Z. Hu, Z. Ye (Eds.), *The Second International Workshop on Education Technology and Computer Science* (pp. 273-276). Los Alamitos, CA: IEEE Computer Society.

Zhang, G., & Liu, W. (2010). The applied research of cloud computing platform architecture in the E-Learning area. In V. Mahadevan & Z. Jianhong (Eds.), *The 2<sup>nd</sup> International Conference on Computer and Automation Engineering* (pp. 356-359). Los Alamitos, CA: IEEE Computer Society.

Zhang, Y., & Chen, J.-L. (2011). A delegation solution for universal identity management in SOA. *IEEE Transactions on Services Computing*, 4(1), 70-81.

## Introducing 1 to 1 in the Classroom: A Large-scale Experience in Chile

Magdalena Claro<sup>1\*</sup>, Miguel Nussbaum<sup>1</sup>, Ximena López<sup>2</sup> and Anita Díaz<sup>2</sup>

<sup>1</sup>Escuela de Ingeniería, Universidad Católica de Chile, Vicuña Mackenna 4860, Macul, Santiago Chile // <sup>2</sup>Eduinnova, Universidad Católica de Chile, Vicuña Mackenna 4860, Macul, Santiago // magdalena.claro@gmail.com, mclarot@uc.cl // mn@ing.puc.cl // ximelopez@gmail.com // adiaz@eduinnova.com

\*Corresponding author

(Submitted January 01, 2012; Revised August 17, 2012; Accepted September 14, 2012)

### ABSTRACT

This paper presents the results of a study that sought to understand how a strategy of Mobile Computer Labs (MCL) has been integrated into 3<sup>rd</sup> and 4<sup>th</sup> grade teaching-learning practices in 1,591 state schools in Chile. In particular, the study aims to identify the most important factors related to MCL integration into these practices a year after its implementation. The findings obtained by applying a survey and performing classroom observations show that, although the schools are willing to use this resource, it has been used sporadically for searching information and drilling and practicing contents related to Mathematics and Language, in addition to motivating students. Furthermore, the classroom observations do not reveal any innovative teaching strategies, related to the use of this new technology. The study shows that amongst the main reasons for this traditional and sporadic use of the MCL are a lack of targeted teacher training and preparation time, and insufficient technical and pedagogical support during the phases of implementation and integration to the pedagogical practices.

### Keywords

ICT in the classroom, One-to-one initiatives, Teaching practices with ICT, ICT strategy evaluation, Teacher support

### Introduction

The Chilean ICT in education policy (Enlaces) started in 1993 and has equipped schools with computers, local networks, educational resources, productivity software, free or subsidized Internet access, as well as providing technical and pedagogical support in partnership with 24 universities from across the country. Computers have been placed in special rooms called Computer Labs and, since 2007 classrooms have been equipped with a multimedia package that includes a personal computer, audio equipment and a projector. By the year 2010 90% of students going to publicly financed schools had access to computers, 60% of schools had Internet access, 110.000 teachers (there are around 180.000 teachers working today in the system) had been trained to use computers as a part of their instruction process, and the Country reached a national average of 9,8 students per computer in schools (Donoso, 2010; Hepp et al., 2004; Sánchez & Salinas, 2008).

For the past few years, the primary concern in Chile has been investigating the effects of these policies. In particular, studies have looked at how these new resources are being used and whether or not there has been any impact on student learning. In Chile, the evidence is consistent with other international studies and reports which show that the frequency of ICT use in teaching and learning activities in schools is relatively low (Cuban, 2001; Hinojosa, et al., 2005; Plomp & Voogt., 2009; OECD, 2010) and that they are mostly used to support “traditional” teaching practices (i.e., instruction based lessons) (Plomp & Voogt, 2009; Trucano, 2005; Balanskat et al., 2006). In fact, data collected through a 2009 national survey of all state-subsidized schools and a sample of private schools showed that ICT is not frequently used in teaching and learning in the classroom. When used, it is to support teacher’s current practices instead of changing or revolutionizing them (Hinojosa et al., 2010).

In terms of the impact on student learning, international and national research has not been able to provide conclusive statements about their positive or negative effects. On the contrary, findings are most often inconsistent between studies or difficult to generalize. These studies are commonly country-specific or developed under particular conditions; such is the case with pilot projects or case studies. Additionally, there are very few experimental studies that allow for empirically sound conclusions related to causality between ICT and student performance (Balanskat et al., 2006; Kozma, 2006; Ungerleider & Burns, 2003; McFarlane et al., 2000; Cuban & Kirkpatrick, 1998). Where evidence seems to converge is in indicating a non-linear and complex relationship between ICT and learning.

The lack of impact of ICT in schools has been explained in different ways by educational researchers and experts. One such explanation is related to the inability of technologies and technological school settings to adapt to real educational needs (Means, 2000; Watson, 2001). As previously mentioned, in Chile ICT resources are found mainly in computer labs, which implies that activities built around these technologies change the natural context of classroom teaching and tend to focus on the purely technological aspects. In this sense, the technologies “(...) are not truly integrated into the classroom teaching dynamic, which may limit their impact on teaching styles traditionally used in schools” (Nussbaum et al., 2009, p. 295). As Watson (2001) argues, having to book a timetabled fixed resource and moving the class to a separate room for a limited time does not allow for open-ended exploratory work, which the technology could facilitate. Software is increasingly seen as having the potential to support and enhance curriculum initiatives based on a conceptual understanding and the development of process skills. Nevertheless, hardware’s physical location and management threaten this potential benefit. Consequently, from this point of view, as technologies adapt to real educational needs, ICT should tend to be used more appropriately and achieve the desired learning results (Means, 2000).

Among the technologies available, interactive whiteboards and laptops, provided on a one-to-one (1:1) basis, have been introduced in the classroom. With regard to laptops it is argued that ubiquitous or 1:1 computing environments may enhance learning. This is because they can provide all students and teachers with continuous access to digital teaching resources within the same classroom dynamics, something that computer labs outside the classroom do not allow for (Bonifaz & Zucher, 2004). Furthermore, 1:1 models of ICT use in education are promoted with the idea that they can provide personalized and student-centered experiences to students within schools and beyond (Severin & Capota, 2011). For example, the One Laptop per Child (OLPC) initiative promised to transform education for world disadvantaged students by giving them the means to teach themselves and each other (Kenneth et al., 2009). The OLPC project was strongly shaped by the visions of the founder of the Massachusetts Institute of Technology (MIT) Media Lab, Nicholas Negroponte and Seymour Papert’s constructionist learning theory. From Papert’s perspective students learning depends on constructing ideas and individual laptop computers can be essential for carrying out such construction in today’s world. Through the OLPC program, Negroponte, Papert and others have sought to develop and distribute a low-cost “children’s machine” that could empower youth to learn without, or in spite of, their schools and teachers (Warschauer & Ames, 2010; Kenneth, 2009).

However, recent studies do not show clear evidence regarding the benefits of 1:1 models of computer use in schools and homes (Warschauer & Ames, 2010). In Latin America and the Caribbean, the Inter-American Development Bank (IDB) recently published a study that concluded that there are still uncertainties relating to the impact of 1:1 programs and that further evaluation is needed (Chong, 2011). The evidence collected in this study and others, so far indicates that programs that overlook teacher training and the development of specific software may yield very low returns (Chong, 2011; Karsenti & Collin, 2011; Severin & Capota, 2011). Additionally, evidence related to projects that give students the opportunity to take their computers home is not positive, especially when it comes to poor students. In fact, Chong et al (2011) found that children with weak adult supervision at home may not spend their time using computers for homework and studying, and therefore this may have no positive impact in their educational achievement. Some studies have even found that access to computers at home may have a negative impact on academic achievement (Vigdor & Ladd, 2010; Malamud & Pop-Eleches, 2010).

In summary, studies that look at ICT use in schools in Chile and around the world show that there is still some way to go before we really understand how to naturally integrate technologies into the teaching and learning process in the classroom and, subsequently, impact learning. Although significant efforts have been made to use technologies that can adapt to natural teaching conditions in the classroom, the benefits are still not clear.

In 2009, Enlaces launched a Mobile Computer Labs (MCL) strategy with the main purpose of developing the capacity of third grade students to read, write and perform basic mathematical operations through the integration of computer equipment. This equipment allows 1:1 learning strategies to be developed. This strategy consisted in providing 1,591 state primary schools with a cabinet of netbooks, one netbook for each student in each third grade classroom and one for the teacher with software that allows them to control and communicate with the class. The cabinet makes possible to move the computers between classrooms, as well as offer storage, security and a means for charging the batteries. It also provides integrated wireless networking technology that allows communication between computers and classroom collaboration (Mineduc, 2009). The strategy also includes a web page that provides information about the project, as well as digital educational resources in Mathematics and Language to support class lessons.

In this context, the main purpose of this study was to evaluate how Chilean schools implemented and used the MCL in the classroom and whether this new strategy had any effect in teaching practices. More specifically, our interest was to study how this new strategy of providing 1:1 technology in the classroom had been integrated into the teaching and learning process of Chilean third grade classrooms. More specifically, our purpose was to identify the most important variables associated with key dimensions in the process of implementation and integration of MCLs in schools and classrooms during its first year of implementation. The research questions were driven by the work of Zhao et al. (2002) who identified three key domains that should be considered when trying to understand the processes involved in implementing new technologies:

- The innovator that uses the technology, in this case, the teacher.
- The innovation or project to be developed with the technology, in this case, the pedagogical use of the MCL.
- The context in which the innovation takes place (i.e., technology infrastructure, human infrastructure and organization culture), in this case the school.

In this study we aimed to identify and analyze the different factors related to these three processes that explain a higher or lower level of integration of the MCL in third grade Chilean classroom teaching practices. Consequently, the research questions were:

- In what way and to what extent are teachers using this technology in their teaching practices?
- What have been the aids/facilitators and barriers for integrating these technologies into teaching practices?
- Which factors associated with human, technological and contextual school conditions are more closely related to the integration of MCL into teaching practices?

## Methodology

### Instruments

A quantitative method was used based on the design and application of a self-administered questionnaire completed by the participating schools. The questionnaire was built taking into consideration the main elements of the MCL strategy, as well as the results of interviews with representatives from the Ministry of Education and other school figures (principals and teachers) from participating institutions.

The questionnaire comprised 147 items, organized in four sections or dimensions: (1) personal characteristics of the respondent; (2) teaching and ICT in the school; (3) use and organization of the MCL; and (4) MCL Project. The first dimension asked about the role of the respondent at the school (i.e., teacher, school director, ICT coordinator), his/her age and ICT access and use at home. Dimensions two and three were related to the organizational conditions found after a literature review, i.e. the context for integrating the MCL in the school and the classroom, such as time for teacher preparation (Jones, 2004, Cox et al., 2004), technical and pedagogical support (Kirkland & Sutch, 2009; Law et al., 2008; Trucano, 2005), the school directors' support and leadership (Law et al., 2008) and schools' ICT plans and strategies (Kozma, 2003). Finally, the fourth dimension asked about the expectations, evaluation and future perspectives related to the MCL project at the school. The questionnaire presented different types of questions: 53 multiple-choice questions, 11 multiple answer questions, 82 4-point Likert scale questions and one open-ended question. Originally the items did not belong to predefined scales. In the data analysis phase, scales were created to facilitate analysis and interpretation of data (see Data Analysis section). The design of the questionnaire considered that it could be answered online in approximately 30 minutes.

### Sample design and data collection

The study's sampling frame comprised the 1,591 schools that participated in the MCL project. Stratified random sampling was applied, using the criteria of Region (Chile's first-order administrative division) and 'rurality' to form the required strata. A probabilistic sampling of schools was then applied to each stratum in order to maintain the proportion of schools in each stratum and guarantee the representativeness of the sample. The sample size was calculated using a 95% confidence interval and 5% margin of error. For each of the selected schools, two 'replacement' schools were also chosen randomly. These schools would be successively incorporated in case the original school failed to answer the survey.

A total of 565 schools were contacted in three successive calls. An email with the URL to access the survey was sent to each school. Only one representative from the school could answer the survey (teacher, Principal, Head of Curriculum and Instruction or Head of ICT). This representative had to be someone who had participated in the MCL project. A total of 242 valid survey responses were gathered for analysis. The follow-up after the third call revealed that 70% of non-responses were due to causes not related to the project (e.g., a nationwide student strike).

Unit non-response was treated by post-stratification weighting based on the known information about the population frame (region and rurality) (Holt & Elliot, 1991). Weighting is the most widely used strategy for handling unit non-response in order to reduce non-response bias (Armoogum & Madre, 1998; Little & Vartivarian, 2004), especially in cases where non-response is not related to the phenomena under study. Weighting coefficients were calculated by dividing the expected number of respondents by the actual number of respondents in each stratum. Table 1 shows the details of the composition of the sample and the weighting coefficients.

*Table 1. Sample Characteristics*

Region	Expected Sample Size			Achieved Sample Size			Weighting Coefficients	
	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural
<b>1</b>	1	1	2	1	1	2	1.00	1.00
<b>2</b>	4	0	4	4	0	4	1.00	-
<b>3</b>	5	2	7	5	2	7	1.00	1.00
<b>4</b>	13	10	23	10	5	15	1.30	2.00
<b>5</b>	18	10	28	12	8	20	1.50	1.25
<b>6</b>	17	18	35	11	15	26	1.55	1.20
<b>7</b>	18	24	42	11	22	33	1.64	1.09
<b>8</b>	35	19	54	25	16	41	1.40	1.19
<b>9</b>	20	7	27	18	5	23	1.10	1.40
<b>10</b>	13	11	24	10	7	17	1.30	1.57
<b>11</b>	2	1	3	1	1	2	2.00	1.00
<b>12</b>	1	0	1	1	0	1	1.00	-
<b>13</b>	53	12	65	37	8	45	1.43	1.50
<b>14</b>	3	2	5	3	2	5	1.00	1.00
<b>15</b>	1	0	1	1	0	1	1.00	-
<b>TOTAL</b>	<b>117</b>	<b>204</b>	<b>321</b>	<b>150</b>	<b>92</b>	<b>242</b>		

## Data analysis

Three types of analysis were carried out:

- The data were analyzed using descriptive statistics for each item (analysis of frequency and percentages). Chi-square values were estimated in order to study the possible differences between sub-groups (e.g., differences due to the role in the school; rurality of the school).
- Given the large quantity of items and sub-items in the questionnaire, new variables (scales) were created using an Optimal Scaling procedure (Greenacre, 2007). Scales were created following two criteria: theoretical coherence of the items and an internal consistency of the scale (Cronbach's alfa) above 0.7. A total of 57 items were grouped in 10 different scales. After quantification, different quantitative analyses were done (e.g., averages, correlations, t-tests and ANOVA) on the aggregated variables.
- Finally, a Binary Logistic Regression analysis was carried out to establish the factors associated with a successful implementation of the project. For the Ministry of Education, investing in the MCL project and implementing the national strategy was worthwhile if the technological resources were frequently used to support learning of third graders, and if the technology was being used to promote diversity and innovation in teaching practices. The frequency of use of the MCL in schools and the pedagogic diversity of use of the MCL



were therefore the two scales chosen as dependent variables. The scale for frequency of use was constructed based on the 6 items that captured information about how much the MCL was used in different subjects. The scale for diversity of use was constructed based on the 8 items that referred to the frequency of use of the MCL in different pedagogical activities. Schools were assigned to two groups (higher/lower frequency of use and higher/lower diversity of use) according to the values attained in each scale. So as to find the optimum number of predictive factors for each of the dependent variables (i.e. find the model with the lowest possible number of variables and maximum predictive capacity), it was decided to carry out a logistic regression analysis with a sequential extraction method based on likelihood (*Forward Stepwise Likelihood Ratio*).

## Results

### Characteristics of the survey respondents

A large number (44.4%) of those who answered the survey were the Head of ICT at their school (Table 2).

Table 2. Percentage of answers per position within school

Position	Frequency (N)	Percentage
Teacher	38	15.8
Head of Curriculum and Instruction	40	16.6
Head of ICT	107	44.4
Principal	53	22.0
Other	4	1.2
Total	242	100.0

The distribution for gender was quite even between men and women, albeit with a slightly higher percentage of women (53%). However, if this is analyzed per position or school role, 70% of principals were men and 79% of teachers were women. Additionally, 43% of respondents were between 41 and 55 years of age and the majority had been at the school for more than 2 years. Finally, with regard to access and use of ICT, 97% answered that they used ICT on a daily basis, 97% that they had a computer at home and 90% that they had a wireless Internet connection.

### Perceptions of ICT and the MCL as a school resource

At the level of believes or perceptions of ICT, the great majority said that, in general, ICT have been well accepted and used as a resource to support new pedagogical practices at their school. In effect, the majority of respondents agreed or strongly agreed that ICT are important for their school (96.6%), that ICT has begun to be used for teaching and learning school subjects (94.7%), that all teachers within the school use ICT in their classes (72.5%) and that there is an institution-wide plan for the pedagogical use of ICT (86.3%).

With regard specifically to the MCL, 92.8% said that their school saw it as an opportunity to promote more innovative pedagogical practices, 82.6% as a way to motivate students (i.e., gain their interest and enthusiasm with the class lesson), 77.3% as a means of improving their students' digital literacy levels, 70.1% as an opportunity to have increased ICT infrastructure for teachers and students, 68.8% as a way of supporting the Institutional Educational Project and 63.6% as a means of accelerating the pedagogical integration of ICT.

### Use of the MCL

The great majority (90.5%) of respondents indicated that their school has been using the MCL for a year or more and in third grade this percentage rises to 95.6%, which was the target class level of the project. As shown in Table 3, the respondents said that the MCL was used with regular frequency (the most frequent answer being "often", followed by "sometimes") in mathematics and language, although more frequently in language than mathematics. If the answers 'often' and 'always' are added together, then 59.5% of respondents used the MCL in mathematics, while 70.7% used it in language. This use can be explained by considering that the aim of the project was to use the MCL in these two subjects, although the majority of respondents said that they were also used sometimes or often in other

subjects such as natural science, history & social science and technology. It is interesting to notice that 33.1% reported never using the MCL for the technology subject, which can be partly explained by Enlaces' strategy to teach technology in the ICT laboratory (Table 3).

*Table 3. Frequency of Use of MCL per Subject*

	<b>Mathematics</b>	<b>Language Arts</b>	<b>Natural Science</b>	<b>History &amp; Social Science</b>	<b>Technology</b>	<b>Other subjects</b>
Always	20.9	24.3	9.8	7.7	5.7	4.6
Often	38.6	46.4	30.8	31.0	21.6	17.3
Sometimes	37.0	27.4	43.3	42.7	39.5	46.9
Never	3.5	1.9	16.1	18.7	33.1	31.2
<b>Total</b>	100.0	100.0	100.0	100.0	100.0	100.0

The correlation analysis showed that the per-subject frequency of use of the MCL is not related to general beliefs or perceptions about ICT, nor the expectations that the schools had with regard to the Enlaces project. Also, there was no significant difference in frequency of use between rural and urban schools, with the exception of Language, where urban schools used the MCL with greater frequency ( $X^2_{(1,317)} = 13.00, p < 0.05$ ).

With regard to the activities that the majority of respondents said that teachers performed 'often' or 'always' with the MCL, the main responses were exercising previously acquired concepts and skills, searching for information and motivating the students. Also, the majority said that teachers used the MCL sporadically (sometimes or never) to evaluate learning, carry out research projects, process and analyze data, give instruction-based classes and introduce contents (Table 4).

*Table 4. Types of Use of the MCL*

	<b>Evaluate learning</b>	<b>Exercise contents</b>	<b>Research projects</b>	<b>Search for information</b>	<b>Process and analyze data</b>	<b>Give instruction-based classes</b>	<b>Introduce contents to be developed</b>	<b>Motivate students</b>
Always	7.3	24.2	6.2	20.3	10.9	14.7	12.6	38.9
Often	21.6	43.5	16.7	31.9	24.4	29.7	32.5	37.3
Sometimes	55.3	30.2	44.9	34.0	42.0	44.7	46.8	22.9
Never	15.8	2.2	32.2	13.8	22.7	10.8	8.1	0.9
<b>Total</b>	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

By comparing the types of use by rurality, urban and rural schools made different use of the MCL. Urban schools made greater use of the MCL for giving instruction-based classes, drilling contents and motivating the students, while rural schools used them more to process and analyze data, evaluate learning and carry out research projects. Urban schools made significantly greater use of the MCL for exercising contents ( $t(287) = 2.09, p = 0.03$ ) and giving instruction-based classes ( $t(281) = 2.06, p = 0.04$ ). These data are interesting, as they suggest that use of the MCL in urban schools tends to be more 'traditional', focusing in teachers giving out contents to students. Instead, in rural schools the tendency seems to be using the technology to explore constructivist didactics more focused on promoting active involvement by the students in research activities and the construction of knowledge.

### **School conditions for MCL use**

During the interview stage, two important conditions for using the MCL came up. One of them was that schools needed to have the administrative measures to allocate the necessary time and space for training and preparing classes. Another one was that teachers also had to have technical and pedagogical support in the classroom. Therefore, these two variables were deemed as essential for measuring the conditions for the use of the MCL.

In terms of administrative measures adopted by the schools, the most frequently cited measures, adopted by around three quarters of schools, were the following: appoint a professional to be responsible for the security and care of the

MCL (79.6%), have a calendar, schedule or timetable for the use of the MCL (79.2%), set a minimum for the amount of hours' use per subject (71.1%), and organize a time and space for training teachers so as to improve the use of the MCL in their subject (71.8%). Following these came initiatives related to appointing a professional to support the teachers during the teaching process using the computers (68.6%), allocating time to teachers for preparing computer-based lessons (62.1%) and leaving the use of the MCL up to each teacher's own initiative (60.4%).

However, here the view of the teachers and the principals turned out to be statistically significantly different for several of these measures (Table 5), with the principal's view being more positive than the teacher's for every measure. Although in general principals' perceptions were more positive than teachers' perceptions, they turned to be statistically significant in those questions related to administrative measures and school conditions for MCL use. In the rest of the scales no statistical differences were observed.

*Table 5. Teacher and Principal's Perception of Administrative Measures for Use of the MCL at School*

	<b>Position</b>		<b>X<sup>2</sup></b>
	<b>Teacher</b>	<b>Principal</b>	
Use of each teacher's own initiative	80.0%	48.6%	X <sup>2</sup> = 12.19, p = 0.000
Allocate a minimum amount of time to the teacher for preparing computer-based lessons	45.7%	74.3%	X <sup>2</sup> = 9.76, p = 0.002
Organize a time and space for training teachers so as to improve the use of the MCL in their subject	65.3%	85.3%	X <sup>2</sup> = 6.40, p = 0.011
Appoint a professional to support the teachers during the teaching process using the computers	51.0%	75.4%	X <sup>2</sup> = 7.49, p = 0.006
Appoint a professional to be responsible for the security and care of the MCL	72.9%	87.1%	X <sup>2</sup> = 3.80, p = 0.051

With regard to pedagogical support, the general perception was that it was not very frequent. The majority said that the different school actors either never or only sometimes supported teachers, with the exception of the Head of ICT at the school, where the majority stated that this person often or always offered support (Table 6). With respect to technical support, the responses were very similar.

*Table 6. Frequency of Pedagogical Support for Using the MCL at School*

<b>Pedagogical support</b>	<b>Colleagues</b>	<b>Management</b>	<b>Head of ICT</b>	<b>Head of Curriculum/ Instruction</b>	<b>External support</b>
Always	14.2	15.5	42.4	21.0	6.7
Often	30.4	17.7	22.5	25.7	17.9
Sometimes	43.4	35.2	27.8	33.7	46.0
Never	12.0	31.6	7.3	19.7	29.4
<b>Total</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>

To analyze the aids and barriers of the strategy, the different participants were asked whether or not they agreed with the various statements shown in Table 7. As it can be observed, main barriers were technical and pedagogical support for using ICT during lessons and availability of time for teachers to prepare, develop and implement new activities using ICT. In line with the results reviewed in the above section about perceptions of ICT, the great majority felt that in general there is a willingness in their school to use technology. Additionally, the majority did not see student or teacher ability or teachers' confidence in using ICT as barriers. Finally, an important majority believed that the infrastructure is adequate.

*Table 7. Barriers for Using the MCL at Schools*

<b>Aids / Barriers</b>	<b>% Yes</b>
ICT are considered useful	99.1
Using ICT in teaching-learning is a goal	92.3
Infrastructure for using ICT is adequate	79.0
Technical support for teachers is sufficient	46.0

Pedagogical support for teachers is sufficient	40.8
Teachers have the necessary ICT skills	54.7
Students have the necessary ICT skills	66.3
Teachers have the necessary confidence	63.2
Teachers have the necessary teaching skills	60.1
Teachers have enough time	32.8
There is an overload of projects	44.2

At the end of the survey, the respondents were asked to give an overall evaluation of the different stages of the MCL project. In line with the answers for the previous item, training and support were mainly rated 'normal', while implementation and quality of equipment were rated 'good'. The evaluation by the users of the MCL was not related to the rurality of their school, i.e. there were no significant differences between the evaluation by users at urban schools and rural schools.

The correlation analysis showed that the overall evaluation of the project had a weak correlation with the frequency of use of the MCL, while there was a moderate correlation with some of the factors related to teacher support and ICT confidence. In particular, there was a moderate correlation between the existence of adequate pedagogical and technical support and the presence of teachers within the establishment with a sufficient level of confidence for using the ICT ( $r = 0.45$ ,  $r = 0.39$  and  $r = 0.37$ , respectively).

### Explanatory factors for use of the MCL

From the Binary Logistic Regression two models were obtained based on the two dependent variables, which were deemed to be indicators of a more or less successful implementation. In the first model, where the dependent variable was the frequency of use of the MCL, six significant predictive factors ( $p < 0.05$ ) were found (Table 8). From this analysis it could be concluded that there was a greater possibility of an increased use of the MCL in schools where there was a perception that students had greater ICT skills, where there was better pedagogical support from colleagues, and more class observations. These last three aspects are consistent with the per-item descriptive analysis, where the importance of teachers feeling supported in their use of the MCL was also evident. The other two aspects (better organization and having basic skills) appear to be relatively necessary conditions for any ICT strategy that is implemented in a school. Better technical support appeared to be a not statistically significant factor ( $p > 0.05$ ), although it contributed to the overall model significance.

Table 8. Model for Frequency of Use of the MCL

Predictor	B	B S.E.	Wald	df	P	e <sup>B</sup> (odds ratio)
Level of student ICT skills	1.2	0.393	9.269	1	0.00	3.3
Existence of planning for MCL use	0.97	0.281	12.002	1	0.00	2.64
Pedagogical support for teachers from colleagues	0.78	0.344	5.117	1	0.02	2.17
Technical support for teachers using the MCL	0.64	0.341	3.503	1	0.06	1.89
Observation of teachers using the MCL	0.97	0.445	4.779	1	0.03	2.64
Person in charge gives Technical Support Network (TSN) training	-1.16	0.529	4.813	1	0.03	0.31
Using ICT to teach or learn is a goal of the establishment	-1.97	0.841	5.513	1	0.02	0.13
Constant	2.52	0.871	8.396	1	0.00	12.4
<b>Overall model evaluation</b>						
-2Log Likelihood	117.95					
Model Chi square	91.009					
Cox & Snell R square	0.452					
Nagelkerke R square	0.604					

The second model, where the dependent variable was diversity of types of use, seven significant predictive factors were found ( $p < 0.05$ ) (Table 9). With this analysis it was evident that the variables of greatest weight were the presence of a calendar for use of the MCL, the existence of lesson planning, the fact that a professional is appointed

to support teachers and support from colleagues. It can be concluded from this analysis that there is a greater probability that teachers will use the MCL in diverse teaching practices if in the school there is a calendar for its use, if there is a plan for integrating the MCL into the different subjects and if the teachers are supported by their colleagues and by a technical professional. As such, the involvement in lesson planning training by the school administration appeared to be a factor favoring the use of the MCL in both traditional and constructivist teaching practices. Although contributing to the overall model significance, lesson planning training given by the company which supplies the equipment appeared not to be a significant explanatory factor.

*Table 9. Model for Diversity of Types of Use of the MCL*

<b>Predictor</b>	<b>B</b>	<b>B S.E.</b>	<b>Wald</b>	<b>df</b>	<b>P</b>	<b>eB (odds ratio)</b>
Level of student ICT skills	0.577	0.249	5.365	1	0.02	1.781
Existence of planning	0.601	0.257	5.491	1	0.02	1.825
Have a calendar for MCL use	2.805	0.786	12.726	1	0.00	16.534
Offer specific digital resources for each subject	-1.547	0.622	6.177	1	0.01	0.213
Assign a professional to support the teachers (e.g. Technician)	1.492	0.512	8.497	1	0.00	4.444
Lesson plan training (Company)	1.897	1.097	2.99	1	0.08	6.665
Lesson plan training (Administration)	1.74	0.779	4.984	1	0.03	5.699
Frequency of pedagogical support for teachers by colleagues	1.365	0.37	13.593	1	0.00	3.915
Constant	-2.231	0.92	5.882	1	0.02	0.107
<b>Overall model evaluation</b>						
-2Log Likelihood	130.28					
Model Chi square	98.856					
Cox & Snell R square	0.449					
Nagelkerke R square	0.599					

### **Classroom observations**

Considering that the survey responses were based on the respondents' perceptions of how the MCL had been working at their school, it was decided to observe some schools that reported a frequent use of the MCL. This analysis cannot be considered as representative of all of the schools that showed a greater use of the MCL. However, these observations do allow for a more in-depth investigation of the types of practices that are being carried out, as well as to learn more about the conditions in which they are implemented and the role played by the MCL in the process.

Frequency and diversity of use of the MCL within the school was used as a first selection criterion. Only schools that said they 'always' use the MCL in Language and Mathematics were chosen. In addition, half of the schools selected declared to use the MCL with more traditional teaching methods (e.g. exercising contents, giving instruction-based lessons) and half said that they made frequent use of the MCL for more constructivist activities (e.g., research projects, data analysis). Geographical location of the schools was used as a second selection criterion, due to time constraints. All schools selected were from Santiago (capital of Chile), which is located in the region with the highest concentration of schools that 'always' used the MCL (30.3%).

A total of 11 schools were visited by the observation team, of which five (four urban and one rural) used the MCL with more traditional methods and six (four urban and two rural) with more constructivist methods. In each school a 45-minute Language or Mathematics class was observed. The observation had been previously agreed with both the school and the teacher, meaning that the teacher could plan the use of the MCL. The observations were carried out by two members of the evaluation team following an observation guideline, which allowed them to record information regarding the integration of the MCL into the class lesson, how the teacher handled the technology, the type of activity carried out using the MCL, as well as the student-teacher and student-student dynamics related to the use of the technology.

The analysis of the observations showed that in five of the eleven classes there were technical problems, making it difficult to carry out the activities and to integrate the MCL with the defined pedagogical objectives. In fact, in one case the technical problems made it impossible to develop the planned activity. It was also observed that the majority of teachers did not have the sufficient knowledge to deal with and solve these technical issues on their own.

In four of the eleven observed classes, the teachers had an assistant who offered technical support in how to manage the technology. Two of these cases reported in the survey a constructivist approach and the other two a traditional approach and two of them delivered a lesson in Mathematics and the other two in Language. Therefore having or not an assistant did not seem to be related with the contents or types of use of the technology. In one of these cases the assistant was the schools' ICT coordinator that had planned in advance the class with the teacher, which allowed the teacher to concentrate in the pedagogical process and obtain a better control of the class. In the classes observed where this support did not exist, the occurrence of technical problems affected the normal development of the class, reducing the effective amount of class time by up to 30 minutes of the 90 minutes class. In two of the cases observed, the students had a high level of ability when it came to handling netbooks, being capable of identifying and resolving technical issues.

By comparing the schools which were identified as either more traditional or constructivist in terms of their use of the MCL, the main differences could be seen in the interaction between the students and the teacher. Even when students used the technology individually, the teachers in the more constructivist schools tended to encourage greater interaction between students and with the teacher.

Despite taking into consideration these differences in teaching methods between the schools, the observations showed that schools used the MCL mostly with the aim of motivating the students, and not to implement more innovative teaching practices by using the technology. In only one of the classes the teacher presented a less traditional teaching activity, using a play-based dynamic that required the students to play an active role and generated a different type of interaction between the students and with the teacher. This was the case where the assistant of the class was the ICT coordinator, which allowed for a better flow and continuity of the class. However, even in this case the activity could have been done without the computers, meaning that the technology supported the innovation but did not drive it.

In summary, the analysis of classroom observations indicates that, even in those schools in which the MCL program could be considered successful due to the reported frequency and type of use of the technology, the introduction of the MCL has not driven teaching practices that differ to traditional paper and pencil activities. In all of the classes that were observed, the MCL was used mainly as a resource to motivate the students and not as an essential tool for carrying out activities that would only be made possible by the use of technology. Moreover, the observed pedagogical activities were strongly affected by technical problems, and thus a frequent use of the MCL did not always imply an effective use from a learning perspective. The observations also showed that in schools in which the MCL is being used in less traditional activities, a greater level of interaction among students and the teacher is being fostered. However we do not know if the teacher followed this practice before the technology introduction.

## **Conclusions and discussion**

The objective of this study was to learn how a new strategy of Mobile Computer Labs providing 1:1 technology in the classroom had been integrated into the teaching and learning process of Chilean third grade classrooms, identifying the most important facilitators and barriers in this process during its first year of implementation. The following conclusions are based on the three main research questions of this study.

With respect to the first question, related to the way and extent to which teachers are using MCL in their teaching practices, it can be concluded that the technology has been used sporadically in Mathematics and Language, with greater frequency in Language than in Mathematics. With regard to the types of activity carried out, the most frequent tended to be exercising previously acquired knowledge and skills and searching for information. Also, more than two thirds of the respondents stated that their school always or often used the MCL to motivate their students. When analyzing the types of use according to rurality, it was observed that urban schools tended to use the MCL to carry out more traditional activities, while rural schools to do more constructivist activities, like developing more

research projects. It is unclear as to what could explain this difference between rural and urban schools, but one possible hypothesis is that in rural areas oral culture takes precedence. In this culture, face-to-face communication and interaction between people play a central role and therefore it could be argued that constructivist methods are adopted more naturally as they promote collaboration and encourage students to play a more active role when interacting with their peers.

In relation to the second question regarding the aids and barriers for integrating the MCL into teaching practices, it was found that the barriers to using the MCL included lack of time for planning, project overload and little technical or pedagogical support. Consistently, the logistic regression analysis found that students' ICT skills, planning activities and technical or pedagogical support all facilitated the use of the MCL in Mathematics and Language.

With regard to the third question of this study related to the human, technological and contextual school conditions related to the integration of MCL into teaching practices, first, in terms of the human factors or conditions, it was found that:

- With respect to participants' beliefs or perceptions, the results consistently showed that there was a willingness and acceptance amongst the different users of both ICT in general and of the MCL in particular. It was also found that the frequency of use in subjects had no correlation with beliefs and perceptions about ICT or general expectations for the project.
- However, it was also found that the principals' view was more positive than that of the teachers in relation to the conditions of training and support for integrating ICT into teaching practices. This is an important finding when considering that different views of reality between participants on the same project (specifically between those that run the project and those that implement it) can be an important cause of difficulty when implementing public policies and strategies. This is also relevant considering that other studies show the importance of a shared vision of the innovation for a successful implementation (Kirkland & Sutch, 2009; Law. et al., 2008; Fullan, 2001).

Second, in terms of the technological conditions, these were consistently positively evaluated. The respondents indicated that they valued the MCL as a resource and the availability of the technology in general was not considered an obstacle to its pedagogical integration.

Finally, with respect to the institutional or contextual conditions, the findings were that:

- As far as the administrative measures implemented in the schools, those related to making the MCL resource more readily available were more positively evaluated than those related to the organization of time and space for planning, teacher training and support.
- Nevertheless, the institutional conditions of pedagogical and technical support emerged as insufficient and an essential condition for integrating the MCL into teaching practices. Even more, the classroom observations showed that those teachers that had technical assistance during the class were able to achieve the defined pedagogical objectives of the class while those that did not have technical assistance during the class failed to achieve these. The relevance of permanent technical and pedagogical support is consistent with other studies findings (Trucano, 2005; Jones, 2004; Cox et al., 2004).

The above conclusions show that the measures taken until now by the MCL project have been insufficient and that teachers require much clearer and more systematic guidance and orientation in order to use the new resources. This is identified as being the main failure of the project and is in line with other studies that have found that the absence of teacher training and development of specific software severely limit the benefits of technology in the classroom (Chong et al., 2011; Karsenti & Collin, 2011; Severin & Capota, 2011). In future stages of this project, or similar projects, teachers cannot be left alone during the process of adoption and implementation of a new technological resource. The classroom observations clearly showed that having someone who supports the teacher before and during the class is an essential condition for a well-implemented activity with MCL. This suggests the need for teachers to plan and work collaboratively with others to integrate ICT into the pedagogical process. It is important to

consider that new technological resources add complexity to the classroom and imply changes in teaching practices and culture that do not occur overnight and which require time and support in order to be adopted and understood by the participants. Therefore, only by generating adequate conditions under which the adoption process is developed, with at least time for preparation and planning, and well designed technical and pedagogical support, can we expect teachers to acquire the necessary skills and new resources to obtain significant effects on student learning.

Finally, some limitations of this study were that it was based on the perception of one schools' representative that answered the survey and not over objective data. In fact, some differences between the perceptions of these representatives (depending in their role in the school) were observed, therefore to a certain extent the evaluation of the project could be determined by the role of the respondent. In addition, the classroom observations were performed within time constraints that did not allow for a complete in-depth study to give a more detailed description of the variables associated with a more frequent and innovative use of MCL. Finally, this study was performed only after a year of implementation of the MCL project in Chile. It is important to continue following up this strategy considering previous limitations and the lessons obtained in this study.

## Acknowledgements

This work was partially funded by a grant of the Chilean Ministry of Education FONIDE # 511051.

## References

- Armoogum, J., & Madre, J. L. (1998). Weighting or imputations? The example of nonresponses for daily trips in the French NPTS. *Journal of Transportation and Statistics*, 1(3), 53-63.
- Balanskat, A., Blamire, R., & Kefala, S. (2006). *The ICT impact report: A review of studies of ICT impact on schools in Europe*. Retrieved from the European Commission website: <http://ec.europa.eu/education/doc/reports/doc/ictimpact.pdf>
- Bonifaz, A., & Zucker, A. (2004). *Lessons learned about providing laptops for all students*. Retrieved from the NEIR-TEC, Education Development Center website: <http://perkinselementary.pbworks.com/f/LaptopLessonsRprt.pdf>
- Cox, M. J., & Webb, M. E. (Eds) (2004). *ICT and Pedagogy: A review of the research literature*. London, UK: British Educational Communications and Technology Agency.
- Cuban, L., & Kirkpatrick, H. (1998). Computers make kids smarter: Right? *Technos Quarterly*, 7(2). Retrieved from Agency for Instructional Technology website: [http://www.ait.net/technos/tq\\_07/2cuban.php](http://www.ait.net/technos/tq_07/2cuban.php)
- Cuban, L. (2001). *Oversold and underused. Computers in the classroom*. Cambridge, MA: Harvard University Press.
- Chong, A. (Ed.) (2011). *Development connections: unveiling the impact of new Information Technologies*. Washington, DC: International Development Bank.
- Donoso, G. (2010). Chapter 8: Enlaces in the Chilean educational system: numbers evolution (Capítulo 8: Enlaces en el sistema escolar chileno: evolución de sus cifras). In A. Bilbao & A. Salinas (Eds.). *The open book of computers in education. Lessons and challenges of Enlaces network* ( El libro abierto de la informática educativa. Lecciones y desafíos de la Red Enlaces). Santiago, Chile: Enlaces, Centre for Education and Technology (Centro de Educación y Tecnología) of the Ministry of Education. .
- Fullan, M. (2001). *The new meaning of educational change*. New York, NY: Teachers College Press.
- Greenacre, M. (2007). *Correspondence analysis in practice* (2nd Ed.). Boca Raton, FL: Taylor & Francis Group.
- Hepp, P., Hinostroza, J. E., Laval, E., & Rehbein, L. (2004). *Technology in Schools: Education, ICT and the knowledge society*. Retrieved July 29, 2012 from: [http://siteresources.worldbank.org/EDUCATION/Resources/278200-1099079877269/547664-1099079947580/ICT\\_report\\_oct04a.pdf](http://siteresources.worldbank.org/EDUCATION/Resources/278200-1099079877269/547664-1099079947580/ICT_report_oct04a.pdf)
- Hinostroza, J. E., Labbé, C., & Claro, M. (2005). ICT in Chilean schools: Students' and teachers' access to and use of ICT. *Human Technology*, 1(2), pp. 246-264.
- Hinostroza, J. E., Labbé, C., Brun, M., & Matamala, C. (2010). Teaching and learning activities in Chilean classrooms: Is ICT making a difference? *Computers & Education*, 57(1), 1358-1367.



- Holt, D., & Elliot, D. (1991). Methods for weighing for unit non-response. *The Statistician*, 40, 333-342.
- Jones, Andrew (2004). *A review of the research literature on barriers to the uptake of ICT by teachers*. Coventry, UK: Becta ICT Research.
- Karsenti, T., & Collin, S. (2011). *Benefits and challenges of using laptops in primary and secondary schools: An investigation at the Eastern Township Schoolboard*. Montreal, Canada: Université de Montreal.
- Kenneth, L., Kraemer, K, Dedrick, J., & Sharma, P. (2009). One laptop per child: Vision versus reality. *Communications of the ACM*, 52(6), 66-73.
- Kirkland, K., & Sutch, D. (2009) *Overcoming the barriers to educational innovation. A literature review*. Retrieved from the Futurelab website: [http://archive.futurelab.org.uk/resources/documents/lit\\_reviews/Barriers\\_to\\_Innovation\\_review.pdf](http://archive.futurelab.org.uk/resources/documents/lit_reviews/Barriers_to_Innovation_review.pdf)
- Kozma, R. B (ed.) (2003). *Technology, innovation and educational change. A Global Perspective. A Report of the Second Information Technology in Education Study Module 2*. Eugene, OR: ISTE
- Kozma, R. B. (2006). Monitoring and Evaluation of ICT for education impact: A review. In D. Wagner, R. Day, T. James, R. Kozma, J. Miller, & T. Unwin (Eds.), *Monitoring and Evaluation of ICT in Education Projects. A Handbook for Developing Countries* (pp. 19-33). Washington, DC: The International Bank for Reconstruction and Development.
- Law, N., Pelgrum, W., & Plomp, T. (Eds.) (2008). *Pedagogy and ICT use in schools around the world: Findings from the IEA SITES 2006 study*. Hong Kong: CERC-Springer.
- Little, R., & Vartivarian, S. (2004). *Does weighting for nonresponse increase the variance of survey means? The University of Michigan Department of Biostatistics Working Paper Series* (Working Paper No. 35). Retrieved September 26, 2011 from <http://www.bepress.com/umichbiostat/paper35>
- Malamud, O., & Pop-Eleches, C (2010). *Home computer use and the development of human capital* (NBER Working Papers No. 15814). Retrieved from the National Bureau of Economic Research website: <http://www.nber.org/papers/w15814.pdf>
- McFarlane, A., Harrison, C., Somekh, B., Scrimshaw, P., & Lewin, C. (2000). *Establishing the relationship between networked technology and attainment: Preliminary study 1*. Coventry, UK: British Educational Communications and Technology Agency.
- Means, B. (2000). Technology in America's schools: Before and after Y2K. In R. S. Brandt (Ed.). *Education in a new era* (pp. 185 -210). Alexandria, VA: ACSD yearbook.
- Mineduc, Enlaces (2009). *Theoric support for the design of the MCL implmentation strategy strategy* (Sustento teórico de las bases para el diseño de la estrategia de implementación de LMC). Education and Technology Center of Chile. Internal document describing the Mobile Computers Lab Strategy. Santiago, Chile: Enlaces, Education and Technology Center of Chile.
- Nussbaum, M., Gomez, F., Mena, J., Imbarack, P., Torres, A., Singer, M., & Mora, M. E. (2009). Technology-supported face-to-face small-group collaborative formative assessment and its integration in the classroom. In D. D. Preiss & R. J. Sternberg (Eds.), *Innovations in Educational Psychology: Perspectives on Learning, Teaching and Human Development* (pp. 295-324). New York, NY: Springer.
- OECD (2010). *Are New Millennium Learners making the grade?* Paris, France: OECD.
- Plomp, T., & Voogt, J (2009). Pedagogical practices and ICT use around the world: Findings from the IEA international comparative study SITES2006. *Education and Information Technologies*, 14(4), 285-292
- Sánchez, J., & Salinas, A. (2008). ICT & learning in Chilean schools: Lessons learned. *Computers & Education*, 51(4), 1621–1633
- Severin, E., & Capota, Ch. (2011). *One-to-One laptop programs in Latin America and the Caribbean* (Technical Notes No. IDB-TN-261). Retrieved July 25, 2012 from [http://wiki.immuexa.com/download/attachments/8048/IDBDOCS-%2335989594-v2-One-to-One\\_Laptop\\_Programs\\_in\\_Latin\\_America\\_and\\_the\\_Caribbean\\_.PDF](http://wiki.immuexa.com/download/attachments/8048/IDBDOCS-%2335989594-v2-One-to-One_Laptop_Programs_in_Latin_America_and_the_Caribbean_.PDF)
- Trucano, M. (2005). *Knowledge maps: ICT in education*. Washington, DC: InfoDev. Retrieved January 20, 2012 from <http://www.infodev.org/en/Publication.8.html>
- Ungerleider, C., & Burns, T. (2003). A systematic review of the effectiveness and efficiency of networked ICT in education. *A state of the art report to the Council of Ministers of Education Canada and Industry Canada* [Unpublished manuscript]. Retrieved from <http://www.cmec.ca/Publications/Lists/Publications/Attachments/55/SystematicReview2003.en.pdf>
- Vigdor, J., & Ladd, H. (2010). *Scaling the digital divide. Home computer technology and student achievement* (Working Paper No. 48). Retrieved from the National Center for Analysis of Longitudinal Data in Education Research website: [www.caldercenter.org/upload/CALDERWorkingPaper\\_48.pdf](http://www.caldercenter.org/upload/CALDERWorkingPaper_48.pdf)

Warschauer M., & Ames, M. (2010). Can one laptop per child save the worlds' poor? *Journal of International Affairs*, 64(1), 33-35.

Watson, D. (2001). Pedagogy before technology: Re-thinking the relationship between ICT and teaching. *Education and Information Technologies*, 6(4), 251–266.

Zhao, Y., Pugh, K., Sheldon, S., & Byers, J. L. (2002). Conditions for Classroom Technology Innovations. *Teachers College Record*, 104(3), 482-515.

## Analysis of Students' Performances during Lab Sessions of Computer Networks Course

**Bulent Gursel Emiroglu\* and Seda Sahin**

Baskent University, Engineering Faculty, Computer Engineering Department, Ankara, Turkey //  
bulentgursel@gmail.com // sahins@baskent.edu.tr

\*Corresponding author

(Submitted May 21, 2012; Revised July 18, 2012; Accepted October 05, 2012)

### ABSTRACT

Due to changes and developments on the information and communication technologies, computer networks gained a critical role for establishing connections to share resources in a specific domain. Sharing information is very important in the digital age for giving high-quality services to all interested people enrolled in the networks. Education of computer networks is an engineering discipline to teach network topics and subjects within the university curriculum. Network courses are offered to the students of the related departments in universities and higher education organizations. In this paper, we conducted a study for analyzing the performances of the computer engineering students at Baskent University, Ankara, Turkey on the Network Course, in both theoretical and practical means. We analyzed the weekly scores collected from the lab sessions throughout the academic semesters of 2007-2008, 2008-2009 and 2009-2010 by using One-way and Two-way Anova and Chi-Square tests. Additionally, students' opinions about the course and lab sessions were evaluated by questionnaires at the end of each semester during 3 years.

### Keywords

Computer networks course, Engineering Education, Applets

### Introduction

Computer network is an infrastructure including computers and peripherals connected to each other for sharing resources. Due to changes and developments on the information and communication technologies, computer networks gained a critical role for establishing connections to share resources in a specific domain. Sharing information is very important in the digital age for giving high-quality services to all interested people enrolled in the networks.

Computer networking offers a basic connection constructed to provide transmission of signals over a distance for the purpose of communication. Computer networks still evolving and the study of computer networks is a necessary part of the computer engineering field. According to Surma (2003), computer network courses are in the main research fields of computer engineering field. He studied a subject based on the students activities which supports course requirements, projects and lab experiments which collaborative and discovery learning techniques were used.

For using computer network applications for educational purposes at universities, Elkateeb & Awad (1999) offered that the usage of multimedia in the educational system increases the quality and reduces the cost. For this reason they produced the Multimedia Learning Center (MLC) learning system and it was used to teach data communication and networking concepts with support of the Web. Similarly, Delialioglu & Yildirim (2007) investigated students' perceptions of the "effective dimensions of interactive learning" in a hybrid course including 25 students enrolled in "Computer Networks and Communication". Moreover, Chiu & Liu (2008) proposed a Network Tutorial System (NTS) to investigate the disadvantages of ineffective CTS (Class Tutorial System) used in Taiwanese universities/colleges and discussed the feasibility of the NTS.

Using simulations for the design and implementation of computer networks is very common due to its availability and flexibility. Bello, Mirabella & Raucea (2007) used a modular web-based test bed comprising basic components (hosts, routers, and access points) implemented on nodes equipped with operating systems and open source software for enabling students to accomplish practical applications of computer networks. Kayssi, El-Hajj, El Asir, & Sayyid (1999) used a Web interface for computer networks course to integrate material, registration, exams and grading.

Moreover, Goyal, Lai, Jain, & Duresi (1998) produced the software package for computer networking and telecommunications. It was a simulation environment with laboratory exercises including a range of network subjects from data message flows to IP (Internet Protocol) fragmentation and reassembly. Similarly, Powell, Anderson, Spiegel, & Pope (2002) defined that Web-based teaching tools can be used to improve the quality of an undergraduate laboratory and to reduce cost with the integration of web-based teaching tools and hands-on laboratory courses. Furthermore, Aziz, El-Sayed, Esche, & Chassapis (2009) applied Web-based educational tools to provide a detailed laboratory experience based on remote and virtual laboratory experiments for undergraduate engineering students.

Rahman, Pakštas, & Wang (2009) conducted a research on network modelling and simulation tools. TinkerNet (Winters, Ausanka-Crues, Kegel, Shimshock, Turner, & Erlinger, 2006) was developed as a low-cost platform for teaching bottom-up, hands-on networking at the undergraduate level. It enables students to build their own networking stack from Ethernet up to TCP (Transmission Control Protocol) or UDP (User Datagram Protocol). Bote-Lorenzo, Asensio-Perez, Gomez-Sanchez, Vega-Gorgojo, & Alario-Hoyos (2010) presented a Distributed Network Simulation Environment (DNSE) based on grid technology used by students as an application.

Sang (2010) presented hands-on laboratory experiments for students by using low-cost Small Office / Home Office (SOHO) networking technologies. Kayssi, Sharafeddine, & Karaki (2004) presented 10 laboratory experiments applied on UNIX workstations for simulating network software.

On the other hand, Anido, Liasas, & Fernaández (2000) presented “learning-by doing” paradigm in the field of virtual laboratories with standalone simulators for distributed virtual environments. Wolf (2010) defined the impact of a virtual laboratory on student learning. Additionally, Jovanovic, Popovic, Markovic, & Jovanovic (2010) presented a Web-based laboratory providing basic principles of computer network. Dobrilovic, Brtko, Berkovic, & Odadzic (2012) offered an approach for the evaluation of the virtual network laboratory and exercises based on the rough set theory for computer-networking course.

On the other hand, IEEE (Institute of Electrical and Electronics Engineers) Computer Society (David, 2004) defines standards for Computer Engineering curriculum including the architectures, protocols, standards and technologies of computer networks. Similarly, ACM (Association for Computing Machinery) (Barry, 2008) defines standards for Information Technology curriculum with the contents of basics of networking.

When we look at the current situation at the universities in Turkey, various applications exist for the network courses. In this part, we analyzed the network courses given in the Computer Engineering Departments of the high-ranked universities of Turkey. For example, in METU (Middle East Technical University), uses of computer networks, network hardware, network software, reference models, networks and data communication services, standardizations, OSI layers are covered (METU, 2010). Similarly, in Boğaziçi University, overview of computer networks, network hardware/software, OSI and TCP/IP reference models, standardization, OSI layers are covered (Boğaziçi University, 2010). However, in Bilkent University, introduction to computer networks and the internet, socket programming, client/server model, peer-to-peer networking, protocols, congestion control, internet routing, OSI layers, local area networks (LAN) and wireless LANs are studied (Bilkent University, 2010). But, all those universities don’t conduct lab sessions for the applications of the theoretical subjects of the network course in a laboratory environment to establish interaction.

When we look at the universities in the United States of America; for example, in Columbia University, protocol layers and encapsulation, IP, connection-oriented and connectionless flows, socket programming, DNS (Domain Name System) / HTTP (Hyper Text Transfer Protocol), reliable data transfer, selective repeat / go-back-N / parity & network coding techniques, flow and congestion control, TCP, switching, routing, bit error detection / correction techniques and MAC topics are covered (Columbia University, 2010). Similarly, in Northwestern University, overview of network architectures, applications (HTTP, FTP), network programming interfaces, transport (TCP, UDP), flow control, congestion control, IP, routing, IPv6, multicast, data link protocols, error-detection/correction, multiple access, LAN, Ethernet, wireless networks, and network security are taught (Northwestern University, 2010). Moreover, in University of California San Diego, principles of computer communication networks with examples from existing architectures, protocols and standards, layering and the OSI model, switching, LAN, MAN and WAN, datagrams and virtual circuits, routing and congestion control, internetworking exists within the course (University of California, San Diego (UCSD), 2010). Furthermore, in University of Washington, basics of networking,

internetworking, protocol layering, framing, error correction, packet and circuit switching, multi-access protocols (Ethernet), queuing, addressing and forwarding (IP), distance vector and link state routing, reliable transport, congestion control (TCP) and security topics are covered (University of Washington, 2010).

However, at British universities in United Kingdom, for example, University of Bristol Computer Science Department offers Computer Networks course covering channels, latency, bandwidth, entropy, noise, capacity, routing, error detecting, data compression, protocols (University of Bristol, 2010). In University College London, physical and link layers, end-to-end arguments, multi-hop networks, reliable transport, congestion control, intra-domain routing: Distance vector, distance vector pathologies, link state, and inter-domain routing: BGP (Border Gateway Protocol), wireless networks are included within the scope of Communications and Networks course (University College London, 2010).

The rest of this paper is organized as follows: Main study including topics and subjects distributed to weekly schedule of the term, objectives and learning outcomes, comparison of weekly performances, questionnaire and results, discussion and conclusion.

## **Material and methods**

In this paper, we present the analysis of the last 3 years' lab performances of the students enrolled in the "Computer Networks" course in Computer Engineering Department, Baskent University. It is a one semester course given in both undergraduate and graduate levels. Our study focused on the undergraduate program. We analyzed the weekly scores collected from the lab sessions throughout the academic semesters of 2007-2008, 2008-2009 and 2009-2010. In every semester, around 40 final-year computer engineering students are enrolled within the course. The course lasts for 14 weeks, 5 hours (3 hours for the theoretical part and 2 hours for the lab sessions) per week. 3 quizzes, homework, a project, a midterm and a final exam are used for evaluating the students' performance. Additionally, lab sessions are conducted to teach practical applications of the course content. Labs start at the 4<sup>th</sup> week and conducted bi-weekly. At the end of each lab sessions, students were asked to answer the questions on paper distributed by the lab assistant. The responses of students were used for the evaluation of their performances.

### **Theoretical part of the network course in Baskent University**

We constructed our course outline based on the 7 layers of the OSI Model. Throughout the term, we taught each OSI layer during the successive weeks, as listed below:

- Orientation week: Introduction to network hardware and software, reference models, network standardization and OSI model.
- 1<sup>st</sup> part: Physical layer: Including fundamentals of data communication and transmission medium.
- 2<sup>nd</sup> part: Data link layer: Design issues, error detection/correction, elementary data link protocols and two sub layers; Medium Access Control (MAC) and Logical link control (LLC).
- 3<sup>rd</sup> part: Network layer: Routing algorithms, congestion control, quality of service, internetworking.
- 4<sup>th</sup> part: Transport layer: Transport service and protocols, UDP / TCP, performance issues.
- 5<sup>th</sup> part: Session layer: Managing sessions between end-user application processes (requests and responses), remote procure calls (RPCs), session recovery.
- 6<sup>th</sup> part: Presentation layer: Two sub layers: CASE (Common Application Service Element) and SASE (Specific Application Service Element), serialization, compression, encryption.
- 7<sup>th</sup> part: Application layer: User and system applications, network security, cryptography, Web security, Social/Ethical issues.

### **Applications of the network course in Baskent University**

Computer Networks course is given at the 7<sup>th</sup> semester. Quizzes, lab experiments, a term project, a midterm and a final are used for evaluation. 5 lab experiments were conducted with around 40 students during 2007, 2008 and 2009 fall terms.

During lab sessions, online supplementary web site of the course text book, Computer Networking: A Top-Down Approach Featuring the Internet, 3<sup>rd</sup> Edition, Kurose & Ross by Addison-Wesley. The supplementary web site ([http://wps.aw.com/aw\\_kurose\\_network\\_3/0,9212,1406346,00.html](http://wps.aw.com/aw_kurose_network_3/0,9212,1406346,00.html)) (Kurose & Ross, 2010) was used for accessing Java applets to perform the lab experiments. Except for the first lab experiment taken from the Web site of EMU (Eastern Mediterranean University, 2008) Computer Engineering Department and web page of Sample Mania assignment (2008), all other lab experiments were done online by using applets on this web site. Students used these applets to experience the related subjects interactively.

Attending to the lab sessions was compulsory, so, all students attended the lab sessions during the terms. At the end of lab sessions, to evaluate the students' lab performances on the related subjects, an evaluation test including 5 questions were distributed to the students and they were asked to answer the questions by writing the solutions on the same paper. In the following section, results of the students' performances were presented with class averages of 3 terms. The students' scores were evaluated and recorded by the lab assistant each week. The lab assistant entered the scores into a spreadsheet program (Microsoft Excel) and the averages of the class were calculated and noted.

### First lab session (during the 4<sup>th</sup> and 5<sup>th</sup> week)

**Objectives:** Students will learn concepts of signal spectrum and representation of a periodic signal in the form of a Fourier expansion. They will find out the concept of sampling of a signal and the effect of different sampling frequencies on the quality of the signal recovered after its sampling.

At the first week, students studied concepts of signal spectrum and representation of a periodic signal in the form of a Fourier expansion with a signal analysis program, as shown in Figure1. The concept of sampling of a signal and the effect of different sampling frequencies on the quality of the signal recovered after its sampling with SampleMania applet, as shown in Figure 2.

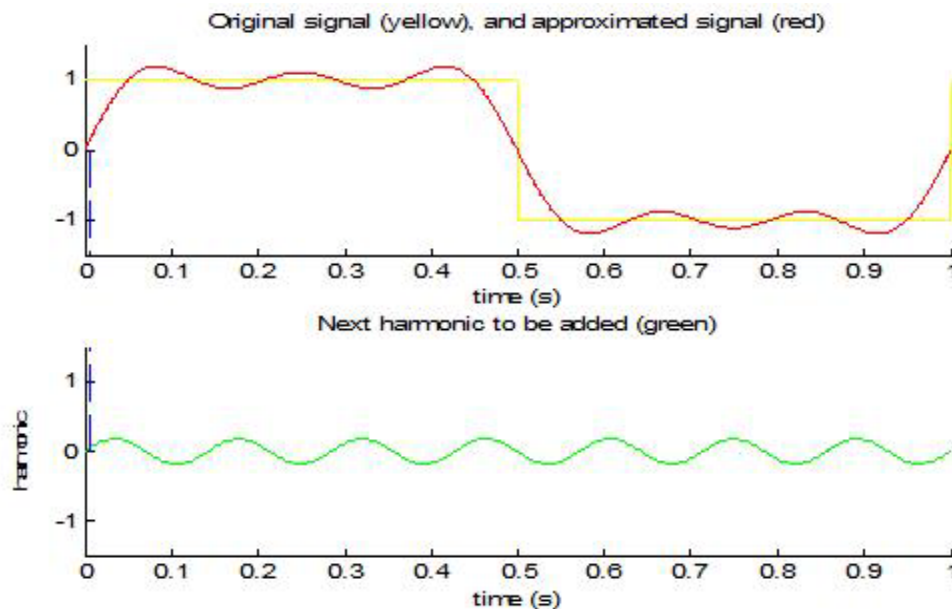


Figure1. The output of Signal Analysis created by Matlab program

**Learning Outcomes:** Students learned the concepts of signal spectrum and the representation of a periodic signal in the form of a Fourier expansion. Moreover, they were informed about sampling of a signal and the effect of different sampling frequencies on the quality of the signal recovered after its sampling.

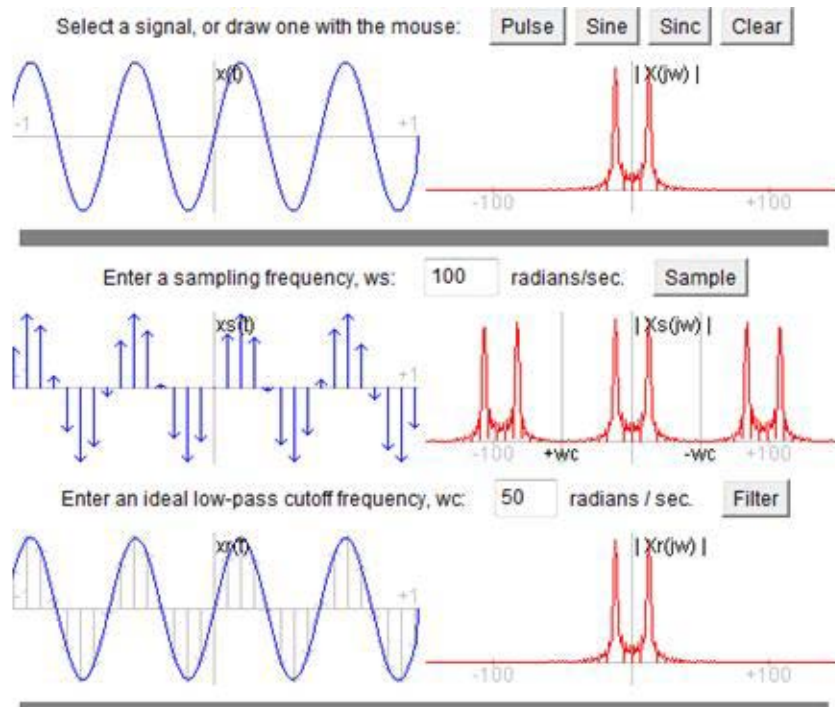


Figure 2. Output of sampling created by SampleMania applet

For measuring the learning outcomes, following questions were asked to the students:

- What is the purpose of this lab work? (20p)
- Draw the data communications model and show the signal transmission over this model and what are definitions of a signal and the signal types? (20p)
- What is the meaning of the Fourier expansion of a periodic signal and what is the meaning of Fourier coefficients in the Fourier expansion of a periodic signal? (20p)
- What is the meaning of the sampling theorem? (20p)
- What is the meaning of the channel capacity? Explain the data rate, bandwidth, noise and error rate concepts and define the maximum channel capacity using the Nyquist and Shannon formulations. (20p)

According to the responses of students, following class averages appeared:

2007-2008 Fall: Class average of 42 students was 69.52

2008-2009 Fall: Class average of 45 students was 71.49

2009-2010 Fall: Class average of 40 students was 73.13

### Second lab session (during the 6<sup>th</sup> and 7<sup>th</sup> week)

Objectives: Students will focus on data communication concepts with the use of simple applets. They will study performance issues including transmission and propagation delays, queuing and loss of data packets and transmission media.

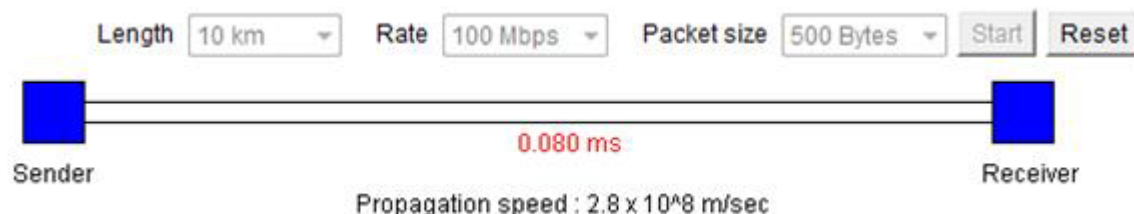


Figure 3. Output of transmission versus propagation applet

In Figure 3, the applet represents the relationship between transmission and propagation time with transmission link length, data rate, and packet size. Students changed values of these terms and created their graphs and discussed these results. For example, in this applet, transmission distance was 10 km, data rate was 100 Mbps and packet size was 500 bytes. However, different variables could give different transmission time results.

After performing the experiments, students reached the following results:

- If transmission link length is decreased, then the transmission time will decrease.
- If data rate is increased, packet transmission time will decrease.
- If packet size is increased, the packet transmission time will increase.
- Propagation delay depends on the transmission distance, so If the distance is increased, the propagation delay will increase and  $\text{Propagation Time} = \text{Transmission distance} / \text{Propagation speed}$ .

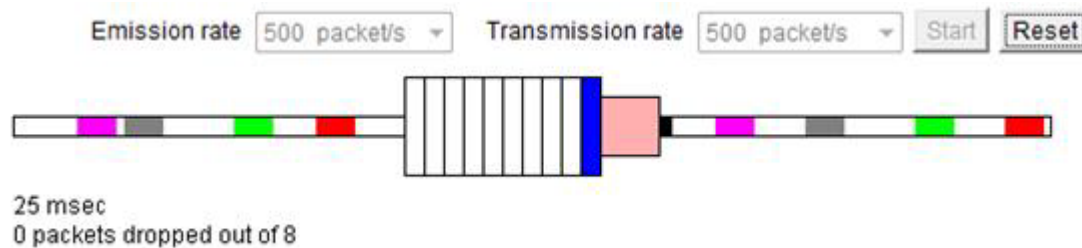


Figure 4. Output of queuing and loss applet

In Figure 4, the applet demonstrates the buffer mechanism for calculating the rate of queuing and loss of data packets. Students tried 3 possible conditions and following results were gathered:

- If emission rate is greater than the transmission rate, then the time duration for packet transmission is not enough, so, buffer may have overflow problem and many packets may get lost.
- If emission rate is equal to the transmission rate, then sending and receiving transmission time rate is adequate for packets, so, no packets get lost.
- If emission rate is smaller than the transmission rate, then we have enough time for sending all of the packets. As a result, the buffer cannot overflow and no packets get lost.

Learning Outcomes: Students learned data communication concepts with the use of applets. They understood performance issues including transmission and propagation delays, queuing and loss of data packets and transmission media.

For measuring the learning outcomes, following questions were asked to the students:

- Explain the purpose of this lab work (20p)
- What are the design factors relating to the transmission medium and the signal which determine the data rate and distance? (20p)
- What is a transmission delay and does the transmission delay depend on the distance between a sender and a receiver? (20p)
- What is a propagation delay and does the propagation delay depend on the used transmission media? (20p)
- What is the purpose of a queue when packets (messages) are being received and why can packets be lost? (20p)

According to the responses of students, following class averages appeared:

2007-2008 Fall: Class average of 42 students was 68.50  
 2008-2009 Fall: Class average of 45 students was 74.16  
 2009-2010 Fall: Class average of 40 students was 75.35

### Third lab session (during the 8<sup>th</sup> and 9<sup>th</sup> week)

Objectives: Students will learn the segmentation of messages and multiplexing of communication channels.



At the third week, segmentation of messages and multiplexing of communication channels were studied, as shown in Figure 5 below:



Figure 5. Output of message segmentation applet

This applet represents message segmentation/switching and packet switching concepts. For message switching, message and packet size had equal values; for packet switching, message size were greater than the packet size. The applet includes four nodes: Source (node A), destination (node B), and two intermediate store-and-forward switches. Each packet sent from node is transmitted over three links before it reaches the destination. Each of these links has a transmission rate of 4 kbps with an optional propagation delay of 1 second and each small rectangle demonstrates 1 Kbit of data. Simulation starts with the grouping of rectangles into 1 packet in the source buffer and the packet is transmitted to the destination. This applet is also a representation for multiplexing concepts.

When students run the simulation, they saw packet switching had smaller end-to-end delay than message switching. Students calculated end-to-end delay and compared their results with those gathered from the applet. They used optional propagation delay in their calculations. Moreover, students learned multiplexing concepts.

Learning Outcome: Students learned segmentation of messages in a network and multiplexing of communication channels. Moreover, students tested end-to-end delay times for message and packet switching separately.

For measuring the learning outcomes, following questions were asked to the students:

- What is the segmentation of a message and why message segmentation is used in a network system? (20p)
- Why do we need some type of switching or multiplexing in communication channels and what are the most often used types of channel multiplexing? (20p)
- What can happen if transmitting and receiving ends of a TDM channel are not synchronized? (20p)
- Three packet-switching networks each contain  $n$  nodes. The first network is a ring has a star topology with central switch, the second has a star topology with central switch and the third is fully interconnected with a wire from every node to every other node. What are the best and worst case transmission paths in hops? (20p)
- Compare the delay in sending an  $m$ -bit message over a  $r$ -hop path in a circuit-switched network and in a packet-switched network. The circuit set up time is  $s$  sec, the propagation delay is  $d$  sec per hop, the packet size is  $p$  bits and the data rate is  $b$  bps. What is the main condition that the packet network has a lower delay? (20p)

According to the responses of students, following class averages appeared:

2007-2008: Class average of 42 students was 70.33  
 2008-2009: Class average of 45 students was 69.84  
 2009-2010: Class average of 40 students was 71.55

#### Fourth lab session (during the 10<sup>th</sup> and 11<sup>th</sup> week)

Objective: Students will learn Go-Back-N and stop-and-wait flow control operation.

At the 4<sup>th</sup> week, students studied Go-Back-N and stop-and-wait flow control operation, as shown in Figure 6 and Figure 7 below:

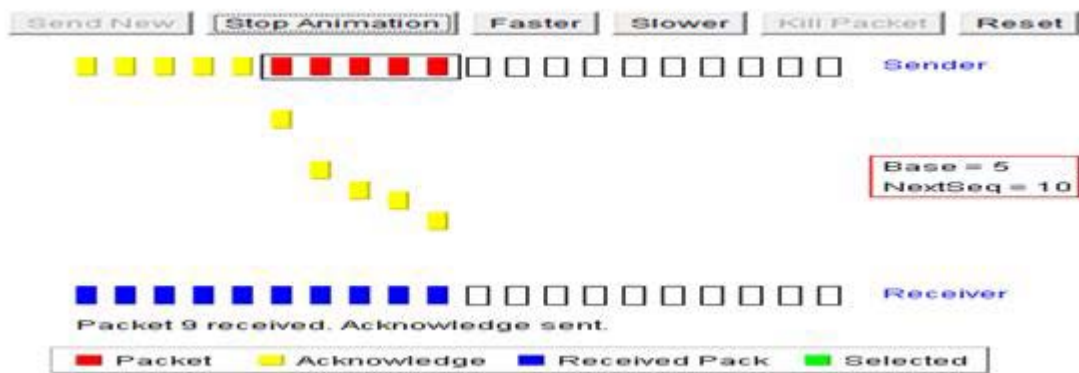


Figure 6. Output of Go-Back-N Protocol Applet

In Figure 6, the applet represents how data link layer protocols and go-back-n sliding window data link layer protocol operate. For testing these, following items accomplished:

- Because the limit of sliding window was 5 here, a maximum of 5 unacknowledged packets could be send to the sender.
- Students sent the packets with the length of sliding window and stopped the simulation before first 5 packets reach the receiver and clicked on the first packet to kill. Then students observed no acknowledgement received from the sender. 5 packets must be sent again and this repetitive transmission of packets was the basic idea for go-back-n protocol.
- Students sent the packets with length of sliding window and stopped the simulation after the first packets reach the receiver and then killed its acknowledgement. Students observed 4 acknowledgements and 1 acknowledgement get lost, which could not cause any problem.
- Students sent more packets comparing to the length of sliding window. Students observed no change occurred and the system permitted to send packets only with the size of the sliding window.



Figure 7. Output of flow control applet

In Figure 7, the applet simulates sending and receiving processes for TCP send/receive buffer. Students tested different combinations of file size and buffer size to learn buffer management for TCP system. They performed necessary computations and interpreted information flow with buffers. In the applet, TCP receive buffer reads with 2Kbyte chunks at random times. If receive buffer becomes full, TCP receiver advertises a receive window of 0 and sender then continues to send segments with 1 byte of data.

Learning Outcomes: Students learned Go-Back-N and stop-and-wait flow control operations. They found out fundamentals of buffer management, sending and receiving processes for TCP send/receive buffers.

For measuring the learning outcomes, following questions were asked to the students:

- Explain the purpose of this lab work. (20p)
- What are the three important mechanisms which are item of data link control?
- What are the differences between stop and wait flow control and the error free sliding window flow control? (Write 3 of them). (20p)
- What is ARQ, piggybacking and pipelining? (20p)
- Sender node (S) and Destination node(D) use a Go-Back-N sliding window Protocol with a window size of 4 and a 3-bit sequence number. If S is a transmitter and D is receiver, show the window positions for the following choices: (20p)
  1. After S sends frames 0, 1, 2 and D acknowledges 0, 1 and the ACKs are received by S.
  2. After S sends frames 3, 4 and 5 and D acknowledges 4 and the ACK is received by S.

According to the responses of students, following class averages appeared:

2007-2008 Fall: Class average of 42 students was 63.74

2008-2009 Fall: Class average of 45 students was 64.18

2009-2010 Fall: Class average of 40 students was 67.78

### Fifth lab session (during the 12<sup>th</sup> and 13<sup>th</sup> week)

Objective: Students will learn CDMA/CD, 802.11 CSMA/CA without hidden terminals and 802.11 CSMA/CA with hidden terminals.

At the 5<sup>th</sup> week, students studied CDMA/CD, 802.11 CSMA/CA without hidden terminals and 802.11 CSMA/CA with hidden terminals, as shown in the applets below:

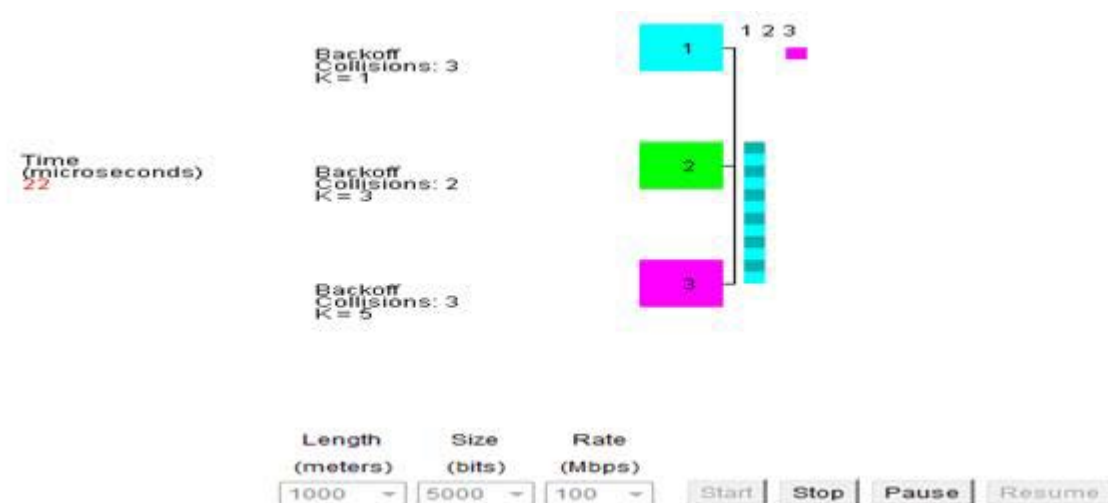


Figure 8. Output of CDMA/CD Applet

In Figure 8, the applet represents relationship between transmission and propagation time with bus length, frame size, and transmission rate for CDMA/CD systems. To do so, students changed the values of these items, created their graphs and discussed the results. Moreover, they analyzed the back off algorithm which is the basic algorithm for both wired and wireless type of CDMA/CD systems.



Figure 9. Output of 802.11 CSMA / CA without hidden terminals applet

In Figure 9, the applet demonstrates a wireless type of CDMA/CD systems called as 802.11 CSMA/CA without Hidden Terminals. There is 1 access point and 3 mobile stations. The mobile stations can listen each other's transmissions. Students analyzed the access point which listens all mobile stations. Each mobile station can listen other mobile stations.

In Figure 10, the applet demonstrates another wireless type of CDMA/CD systems called as 802.11 CSMA/CA systems with hidden terminals. There is 1 access point and 3 mobile stations. The mobile stations can not listen each other's transmissions. Students analyzed the access point which listens none of the mobile stations. No mobile station can listen other mobile stations.

Learning Outcomes: Students learned Carrier Sense on Multi-Access Networks (CSMA) (CSMA/CD and CSMA/CA). They tested relationships between transmission and propagation time with bus length, frame size, and transmission rate for CDMA/CD systems. Moreover, they found out concepts of CDMA/CD, 802.11 CSMA/CA without hidden terminals and 802.11 CSMA/CA with hidden terminals.

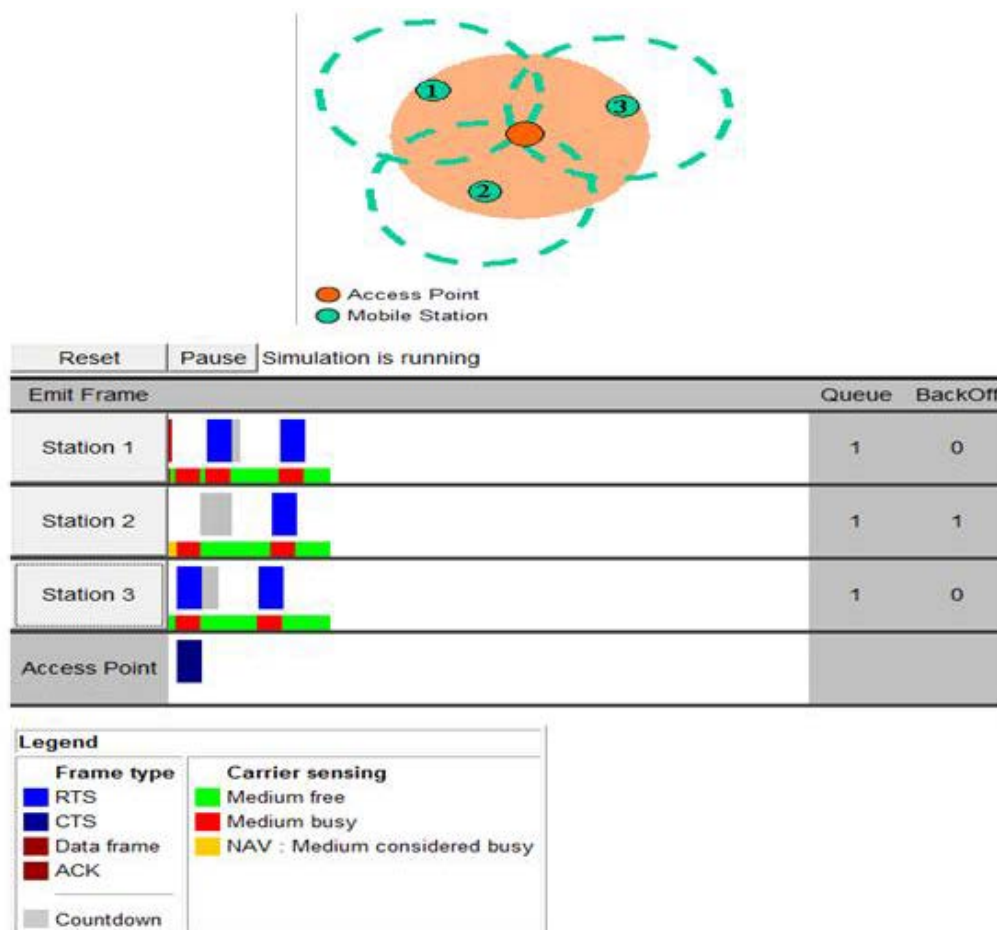


Figure 10. Output of 802.11 CSMA / CA with hidden terminals applet

For measuring the learning outcomes, following questions were asked to the students:

- Explain the purpose of this lab work. (20p)
- What are the LAN technologies and the related MAC techniques which are used by them? (20p)
- If two or more computers happen to choose nearly the same amount of delay after a collision for an Ethernet System, they will both begin to transmit at nearly the same time and producing a second collision. What is the solution to avoid a sequence of collisions? (20p)
- What is the minimum frame size if a CSMA/CD network running at 5 Gbps over a 1-km cable with no repeaters in a building and the signal speed in the cable is 500000 km/sec? (20p)
- What is the most important advantage of CSMA/CD systems over network? (20p)

According to the responses of students, following class averages appeared:

2007-2008 Fall: Class average of 42 students was 71.36

2008-2009 Fall: Class average of 45 students was 72.89

2009-2010 Fall: Class average of 40 students was 76.45

## Results

### Comparison of weekly performances based on 3 groups

As shown in Table 1, when we check the averages of 3 academic semesters, students showed the best lab performance (average is 73.57 over 100) on the 5<sup>th</sup> Lab Session. According to the results, it can be stated that students of 2009-2010 academic semester are the best performing (76.45 over 100 on Lab 5) amongst others.

Table 1. Weekly performance of students based on 3 different academic years

Years	Lab 1	SD	Lab 2	SD	Lab 3	SD	Lab 4	SD	Lab 5	SD
2009-2010	73.13	23.32	75.35	19.30	71.55	21.46	67.78	21.52	76.45	18.21
2008-2009	71.49	23.95	74.16	18.94	69.84	15.20	64.18	13.50	72.89	24.22
2007-2008	69.52	14.47	68.50	19.74	70.33	18.21	63.74	26.05	71.36	24.74
Averages	71.38		72.67		70.57		65.23		73.57	

For the analysis of results gained from the weekly performances of students, One-Way Anova, Two-Way Anova and Chi-Square Tests were applied consequently in Minitab statistical software package.

### Testing for normality

Firstly, normality was tested to determine the normal distribution for our data set and to control suitability of it for other tests. Anderson-Darling test was used for testing normality of lab performances and represented in Figure 11. Since p\_value (0.738) is greater than significance level value of  $p = 0.05$ , our data set was normally distributed.

### Testing for homogeneity

Secondly, homogeneity of variances was tested before Anova tests. Two different tests were applied, and as shown in Figure 12, both of them have greater p\_values, proving that variances of our data set were roughly same.

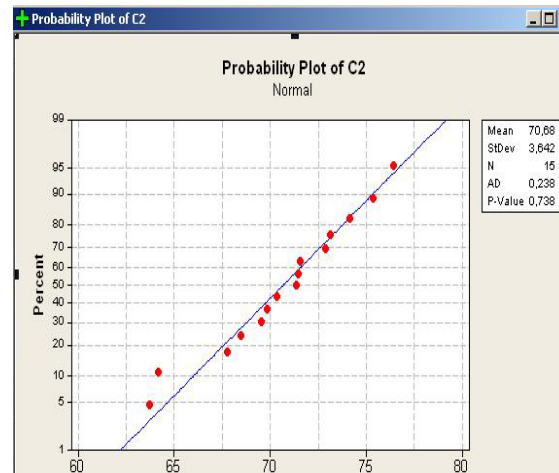


Figure 11. Normality of Lab Performances

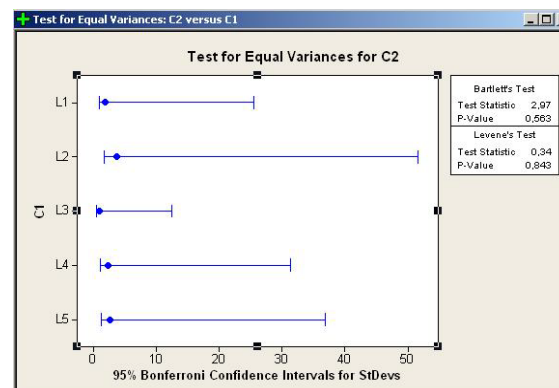


Figure 12. Homogeneity of Lab Performances



## Results for one-way Anova

In One-Way Anova test, obtained p-value (0.014) of the lab sessions is less than the significance level value of  $p = 0.05$ . It shows that at least one score has significant difference. With respect to One-way Anova results of three years' scores, there is at least one significant difference amongst lab sessions, as shown in Figure 13.

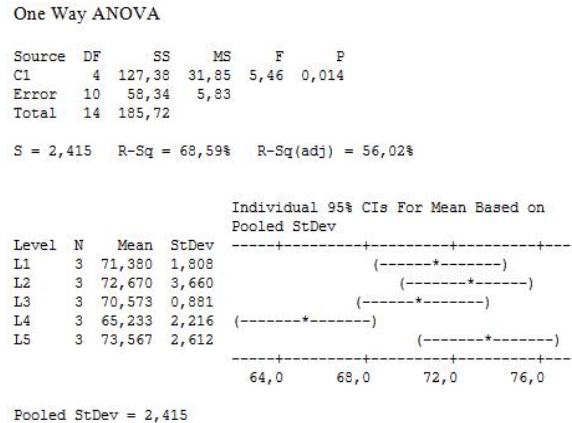


Figure 13. One-Way Anova test

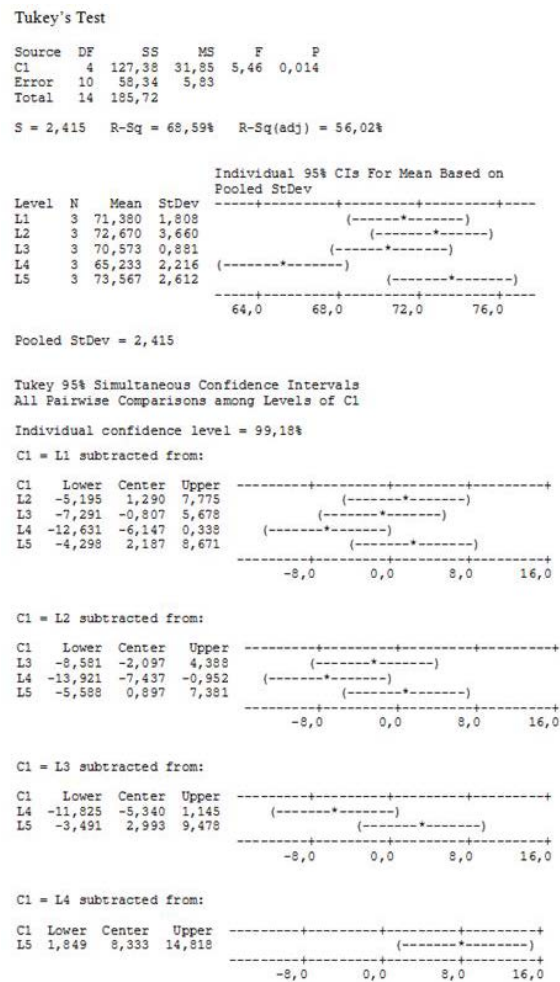


Figure 14. Tukey's test

## Results for two-way Anova

In Two-way Anova test, analysis of variance was performed with two factors, including Lab performances and years. In this test, obtained p-values (0.001 and 0.004 respectively) are less than the significance level value of  $p = 0.05$ . Also, it shows that at least one score is significantly different than others, as shown in Figure 15.

Two-way ANOVA

Source	DF	SS	MS	F	P
C1	4	127,145	31,7863	17,15	0,001
C2	2	43,446	21,7230	11,72	0,004
Error	8	14,825	1,8532		
Total	14	185,416			

S = 1,361 R-Sq = 92,00% R-Sq(adj) = 86,01%

Figure 15. Two-Way Anova test

## Results for chi-square

Chi-square test was applied to compare the intervals for lab scores of students with respect to each academic year and to perform contingency test for our data. As shown in Figure 16, obtained p-value = 0.000 is less than the significance level value of  $p = 0.05$ . This result shows that there is a dependency between the intervals for lab scores of students and academic years.

Chi-Square Test

Expected counts are printed below observed counts  
Chi-Square contributions are printed below expected counts

	C1	C2	C3	C4	C5	Total
1	6 8,19 0,585	16 5,67 18,825	42 47,24 0,582	67 91,65 6,631	69 47,24 10,019	200
2	9 9,21 0,005	2 6,38 3,005	55 53,15 0,064	110 103,11 0,460	49 53,15 0,324	225
3	11 8,60 0,671	0 5,95 5,953	53 49,61 0,232	114 96,24 3,279	32 49,61 6,249	210
Total	26	18	150	291	150	635

Chi-Sq = 56,884; DF = 8; P-Value = 0,000

Figure 16. Chi-square test

According to these results, it can be stated that students of 2009-2010 academic semester are the best performing amongst others. This result can be explained with the scores students got in the university entrance exam. In Turkey, a centralized exam is used for entering the university. Every year, students were selected for the undergraduate programs of the universities by this centrally administered exam. The organization responsible for the administration of this exam is The Student Selection and Placement Center, affiliated to The Higher Education Council (2011).

Table 2. Upper and Lower Bound Scores of Baskent University Computer Engineering Department for the University Entrance Exam

Exam Year	Lower Bound	Upper Bound
2003 June	321	361
2004 June	328	360
2005 June	329	362

As seen in Table 2, the students who entered the university in year 2005 has taken Computer Networks course on their final year during 2009-2010 academic semester.



### Questionnaire for students' opinions about the lab sessions

At the end of each term, paper-based questionnaire were distributed to the students to gather their views and opinions about the online simulations and lab sessions. We used a five-level Likert scale, including the items listed as follows:

- Strongly disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly agree

Within the questionnaire, 10 statements were asked to the students as shown in Table 3. Mean scores and standard deviations (SD) were obtained by the students' responses of the related term. Means and SDs were calculated with different number of students of each term (2007-2008 Fall, 2008-2009 Fall and 2009-2010 Fall).

*Table 3. Students' Responses for the Usefulness of the Online Simulations in the Lab Sessions*

Statements	2007-2008		2008-2009		2009-2010	
	Mean	SD	Mean	SD	Mean	SD
1. I like to use online simulations on the Web during the Lab sessions	3.48	0.50	3.71	0.46	4.60	0.49
2. I like to receive information e-mails from the Instructor and Assistant of the course prior to the lecture.	3.62	0.62	4.33	0.88	4.78	0.42
3. Interacting with the simulations during the Lab sessions help me to learn.	3.52	0.55	3.51	0.50	4.88	0.35
4. I also use online simulations after lab sessions at home.	2.65	0.49	3.67	0.47	3.78	0.97
5. I compare my performance with other students during the lab sessions.	3.12	0.70	4.47	0.62	4.15	0.70
6. Simulations are useful for me to apply theoretical knowledge on practical applications.	3.50	0.80	4.16	0.52	4.43	0.50
7. I also want to design and develop simulations.	2.93	0.26	4.31	0.93	4.33	0.47
8. I wish there will be more simulations about topics of the course content.	3.27	0.67	4.52	0.63	4.95	0.22
9. I wish there will be more lab sessions during the term.	4.40	0.73	4.74	0.58	4.93	0.27
10. Lab sessions were not boring for me, I enjoyed them all.	4.02	0.84	4.87	0.34	4.98	0.16

Mean scores of the students' responses were compared by using statistical software, MiniTab. According to the responses in Table 3, the lowest scores (2.65, 3.67 and 3.68 over 5.00) belong to the item "I also use online simulations after lab sessions at home". It shows that most of the students don't prefer to study at home. To solve this problem, some of the lab exercises can be assigned to the students as homework to increase the students' motivation for studying at home. On the other hand, the highest scores (4.02, 4.87 and 4.98 over 5.00) belong to the items "Lab sessions were not boring for me, I enjoyed them all." It shows that students were happy to be involved in practical applications during lab sessions.

In Table 3, mean scores of the students' responses are increasing from 2007-2008 to 2009-2010, except for the 5th item: "I compare my performance with other students during the lab sessions." The scores of this item decreased from 4.47 to 4.15 from 2008-2009 to 2009-2010 years. According to the questionnaire, it can be stated that students of 2009-2010 academic semester has the best mean scores about online simulations and lab sessions except on the subject about the comparison of student performances with others.

### Discussion

This study covers network applications in educational environment to analyze how real-life situations about computer networks can be simulated for teaching in engineering courses. There is no internationally accepted standard for the curriculum of computer networks course in the undergraduate level of university education. Various universities design their own curriculum with the help of existing instructors, instructional materials, available

computer labs and infrastructure. As a result, every faculty/department conducts network courses differently, leading to irregularity and chaos in this domain.

Moreover, there is no internationally accepted tool/platform for simulating network applications in the lab sessions of the computer networks courses. There are many tools on the Web for demonstrating network conditions to the students in different ways. Students use those tools according to their preferences and just follow the directions given to them by the instructor. As a result, interaction is limited and insufficient for reflecting the creativity of those students.

In our study, we tried to examine the effects of existing tools (applets, animations and simulations) on teaching of computer networks in computer engineering curricula. Analysis of weekly lab performance scores showed us that students learn better with involving in practical applications of computer networks. Applets used for simulating the concepts and subjects during lab sessions. But, the number of applets available is limited to specific topics and not covering all. Students agreed that simulations are useful for them to apply theoretical knowledge on practical applications and they asked for more lab sessions with more simulations provided.

Students enjoy the lab sessions as they involve actively in the learning process but they don't prefer to use and study applets at home after lab sessions. Additionally, some students don't like to compare their performance with others during the lab sessions. Maybe group work can be encouraged to increase the motivation of students for sharing their knowledge and abilities with others. Moreover, a group project can be assigned to the student to encourage them for studying both in lab and at home.

At this point, it can be stated that theoretical parts conducted in classrooms can be combined with the lab sessions by lecturing in the computer labs. If possible, labs specially designed and equipped for applications of computer networks can be used. To do so, constructive methodology can be preferred to enable students benefit from learning by doing paradigm. According to the scores of the lab performances and questionnaire results, it is clear that students learn concepts of computer networks much better by actively using the related applets.

## Conclusions

The main problem for the conduction of computer networks courses is that most of the existing computer labs at the universities are not sufficient for teaching concepts and subjects of the course effectively. As a solution, animations and applets are used during lab sessions for simulating the concepts of computer networks. This study aimed to show the effects of applets on the learning activities of the computer engineering students. We mentioned about the routine operational procedure for Computer Networks course in Baskent University including details of theoretical information given in lectures and practical applications conducted in lab sessions.

Students' performances for weekly lab sessions were analyzed and scores of 2007 - 2008, 2008 - 2009 and 2009-2010 were compared to see the similarities and differences. According to results, students of 2009-2010 year were better than others. This result was explained with the scores students got in the university entrance exam for that year. Additionally, a questionnaire was applied at the end of every year for evaluating the students' satisfaction about animations and applets used for lab sessions. Questionnaire results showed that students were happy to be involved in practical applications of computer networks subjects during lab sessions. However, additional practice needed to encourage students for studying at their own time and place, after the lab sessions.

Results of this study showed that blending practical applications with theoretical parts of any subject domain during the process of higher education of engineering may help for increasing the quality level of the graduates. For training prospective engineers and scientists of computer science in a better way, universities should invest more in hardware and software of computer labs to simulate real cases in a more economical way.

## References

Anido, L., Liams, M., & Fernández, M. J. (2000). Labware for the Internet. *Computer Applications in Engineering Education*, 8(3-4), 201-208.

- Aziz, El-Sayed S., Esche, S. K., & Chassapis, C. (2009). Content-rich interactive online laboratory systems. *Computer Applications in Engineering Education*, 17(1), 61–79.
- Bello, L., Mirabella, O., & Raucea, A. (2007). Design and implementation of an educational testbed for experiencing with industrial communication networks. *IEEE Transactions On Industrial Electronics*, 54(6), 3122–3133.
- Bilkent University (2010). Computer engineering department web site course content of CS421. Retrieved April 10, 2010, from <http://www.cs.bilkent.edu.tr/>
- Boğaziçi University (2010). Computer engineering department web site course content of CMPE475. Retrieved April 10, 2010, from <http://www.cmpe.boun.edu.tr>
- Bote-Lorenzo, M. L., Asensio-Perez, J. I., Gomez-Sanchez, E., Vega-Gorgojo, G., & C. Alario-Hoyos (2010). A grid service-based distributed network simulation environment for computer networks education. *Computer Applications in Engineering Education*, (20)4 doi: 10.1002/cae.20435
- Chiu, J.-M., & Liu, W. L. (2008). A study of the feasibility of network tutorial system in Taiwan. *Educational Technology & Society*, 11 (1), 208–225.
- Columbia University (2010). Computer engineering department web site course content of CSEE W4119. Retrieved June, 19, 2010, from <http://www.compeng.columbia.edu/>
- Delialioğlu, O., & Yildirim, Z. (2007). Students' perceptions on effective dimensions of interactive learning in a blended, learning environment. *Educational Technology & Society*, 10 (2), 133–146.
- Dobrilovic, D., Brtko, V., Berkovic, I., & Odadzic, B. (2012). Evaluation of the virtual network laboratory exercises using a method based on the rough set theory. *Computer Applications in Engineering Education*, 20 (1), 29 – 37.
- Eastern Mediterranean University (2008). Computer engineering department web site course content of CMPE344. Retrieved October 12, 2008, from <http://cmpe.emu.edu.tr/cmpetr/>
- Elkateeb, A., & Awad, A. (1999). A WWW-based multimedia center for learning data communications and networks. *Proceedings of the 1999 IEEE Canadian Conference on Electrical and Computer Engineering* (pp. 202–208). doi: 10.1109/CCECE.1999.807196
- Goyal, R., Lai, S., Jain, R., & Duresi, A. (1998). Laboratories for data communications and computer networks. *Proceedings of the 28<sup>th</sup> Frontiers in Education Conference*, (pp. 1113–1118). doi: 10.1109/FIE.1998.738577
- David, S. (Eds.) (2004). *Computing curriculum—Computer engineering (CE2004) final curriculum report*. KA: IEEE Computer Society.
- Barry, M. L. (Eds) (2008). *2008 Curriculum guidelines for undergraduate degree programs in information technology*. Provo, UT: Association for Computing Machinery.
- Jovanovic, N., Popovic, R., Markovic, S., & Jovanovic, Z. (2010). Web laboratory for computer network. *Computer Applications in Engineering Education*, 20(3), 493–502. doi: 10.1002/cae.20417.
- Kayssi, A., El-Hajj, A., El Assir, M., & Sayyid, R. (1999). Web-based tutoring and testing in a computer networks course. *Computer Applications in Engineering Education*, 7(1), 1–7.
- Kayssi, A., Sharafeddine, S., & Karaki, H. (2004). Computer-based laboratory for data communications and computer networking. *Computer Applications in Engineering Education*, 12(2), 84–97.
- Kurose J. F., & Ross K. W. (2010). *Computer networking: A top-down approach featuring the Internet 3<sup>rd</sup> edition student resources web site*. Retrieved July 19, 2010, from [http://wps.aw.com/aw\\_kurose\\_network\\_3/0,9212,1406346-,00.html](http://wps.aw.com/aw_kurose_network_3/0,9212,1406346-,00.html)
- METU (2010). Computer engineering department web site course content of CENG436. Retrieved April 10, 2010, from <http://www.ceng.metu.edu.tr/>
- Northwestern University (2010). Computer engineering department web site course content of EECS 333. Retrieved June 19, 2010, from <http://www.eecs.northwestern.edu/academics/computerengineering.html>.
- Powell, R. M., Anderson, H., Spiegel J. V., & Pope, D. P. (2002). Using web-based technology in laboratory instruction to reduce costs. *Computer Applications in Engineering Education*, 10(4), 204–214.
- Rahman, M. A. , Pakštas, A., & Wang, F. Z. (2009). Network modelling and simulation tools. *Simulation Modelling Practice and Theory*, 17(6), 1011–1031.

Sample Mania assignment (2008). *Material for Signals and Systems (520.214) lecture*. Department of Electrical and Computer Engineering, Johns Hopkins University. Retrieved October 13, 2008, from <http://www.jhu.edu/~signals>

Sang, J., (2010). Hands-on laboratory experiments with SOHO networking technologies. *Computer Applications in Engineering Education*, 35(2), 7-13.

Surma, D. R. (2003). Lab exercises and learning activities for courses in computer networks. *Proceedings of the 33rd Frontiers in Education Conference* (pp. T2C-21-5). doi: 10.1109/FIE.2003.1263297

The Higher Education (2011). Council of Turkey. Retrieved May 15, 2011, from <http://www.yok.gov.tr/en/>

University College London (2010). Computer science department web site course content of COMP 3035. Retrieved September 15, 2010, from <http://www.cs.ucl.ac.uk/>

University of Bristol (2010). Computer science department web site course content of COMS12500. Retrieved August 12, 2010, from <http://www.cs.bris.ac.uk/>

University of California San Diego (2010). Computer engineering department web site course content of CSE123. Retrieved July 11, 2010, from <http://www-cse.ucsd.edu/>

University of Washington (2010). Computer engineering department web site course content of CSE461. Retrieved July 11, 2010, from <http://www.cs.washington.edu/>

Winters, T., Ausanka-Cruess, R., Kegel, M., Shimshock, E., Turner, D., & Erlinger, M., (2006). TinkerNet. In D. Tolhurst & S. Mann (Eds.), *Proceedings of the 8th Australasian Conference on Computing Education* (pp. 253-259). Darlinghurst, Australia: Australian Computer Society.

Wolf, T. (2010). Assessing student learning in a virtual laboratory environment. *IEEE Transactions On Education*, 53(2), 216-222.

## Classroom-based Cognitive Diagnostic Model For a Teacher-made Fraction-Decimal Test

Tsai-Wei Huang<sup>1\*</sup> and Pei-Chen Wu<sup>2</sup>

<sup>1</sup>National Chiayi University, 85 Wunlong Village, Minsyong, Chiayi 62103, Taiwan // <sup>2</sup>National Pingtung University of Education, 4-18 Minsheng Rd., Pingtung City, Pingtung County 90003, Taiwan // [twhuang@mail.ncyu.edu.tw](mailto:twhuang@mail.ncyu.edu.tw) // [pcwu@mail.npue.edu.tw](mailto:pcwu@mail.npue.edu.tw)

\*Corresponding author

(Submitted January 05, 2012; Revised March 15, 2012; Accepted October 23, 2012)

### ABSTRACT

Teacher-made tests have recently gained attention because of their functions of formative assessment in students' learning. This study aimed at examining the quality of a teacher-made test and diagnosing students' learning misconceptions on fractions and decimals based on a classroom-based cognitive diagnostic BW model that provides four person-facet aberrance indices (capability, misconception, guessing, and carelessness) and four item-facet ones (difficulty, indistinctness, disturbance, and hint). A teacher-made test on the unit of fractions and decimals containing 22 items on five concepts was taken by 32 students in a fourth-grade mathematics class in Taiwan. The results showed that students were classified into four classes with normal, guessing, careless, and complex based on their response patterns. With a good agreement between persons and items, the BW model showed effective examinations for the misfit of items and individuals in the small-sample classroom scenario. The concepts of *equivalent fractions* and *transferring fractions into decimals* were most and least mastered for the current sample, respectively.

### Keywords

Cognitive diagnostic model, BW model, Teacher-made test, Fractions and decimals

### Introduction

Teacher-made tests have recently gained attention because of their functions of formative assessment in students' learning (Dekker, 2007; Frey, Petersen, Edwards, Pedrotti, & Peyton, 2005; Frey & Schmitt, 2010; Holler, Gareis, Martin, Clouser, & Miller, 2008; Nodoushan, 2011; Wiliam, Lee, Harrison, & Black, 2004). To facilitate students' learning effectively, teachers need the competencies of test construction and learning diagnoses in class. Although several pedagogical disciplines have taught teachers how to construct items (Frey et al., 2005; Hopkins, 1998; Kubiszyn & Borich, 2007), how well tests have been conducted was frequently based on subjectively judgmental criteria. Items can be objectively qualified only after a large number of samples had participated on the tests. This, however, seems impractical in classroom setting because no large samples can be provided in a traditionally sized class and even if this is the case, the qualified items might lose their original assessment purposes.

As known, teacher-made tests can fulfill the diagnostic function of understanding students' misconceptions on a certain subject unit. A misconception or an error occurs when expected knowledge by experts or teachers is structurally inconsistent with a student's or novice's actual knowledge (Hartnett & Gelman, 1998). Analyzing students' learning errors to enhance their conceptual understandings has been highly valued and recognized in many fields (e.g., Borasi, 1994; Hartman, 2001). In mathematics, students often find it difficult to learn fractions because they are quite different from whole numbers and may involve many complex quantities (such as decimals) and even language (Bezuk & Bieck, 1993; Mack, 1995; Ni, 2001; Ni & Zhou, 2005). A common misunderstanding related to fractions on multiplication and division is that "multiplication makes bigger" and "division makes smaller." In other words, students expect a greater product and a smaller quotient than both whole numbers themselves (Greer, 1988; Sowder, 1988). To understand students' conceptual misunderstandings, procedural mistakes, or even psychological response aberrances (e.g., guessing and carelessness) in fractions and decimals, teachers undoubtedly need some immediately detailed feedbacks from students' responses on the teacher-made tests that they will use in the classroom.

To analyze what students had mastered or misunderstood, many cognitive diagnosis models (CDMs) were developed to provide informative profiles about students' learning and progresses. Unlike the item response theory (IRT) models only providing item characteristics estimated (e.g., difficulty, discrimination, and guessing parameters), the

CDMs make it possible to investigate underlying content knowledge in items (McGlohen & Chang, 2008). However, as McGlohen and Chang indicated, most CDMs can estimate only students' knowledge states but not their ability levels. In addition, most CDMs are still novel and limited because of a few readily available computer programs (de la Torre, 2009). Moreover, desirable parameters in CDMs are usually estimated through large sample sizes. Consequently, being lack of convenient tools for estimation and big sample size requirements, most teachers in classroom are not acquainted with and even fear using modern CDMs. Fortunately, aberrance indices can be deemed as an adequate option for CDMs because of their in-depth reflections on students' misconceptions without through large samples or complex estimation requirements (Liu, Douglas & Henson, 2009; Liu & Yu, 2011; Seol, 1999).

Typically, aberrance indices may be classified into two areas: IRT-based and group-based categories (Meijer & Sijtsma, 1999). The IRT-based indices usually use a specific IRT model as a standard norm so that any response pattern departing from the model will be classified as aberrant. They may include the residual-based approaches (e.g., OUTFIT/INFIT, Smith, 1991; Linacre & Wright, 1994), the likelihood-based approaches (e.g.,  $l_0$ ; Levine & Rubin, 1979), and the covariance-based approaches (e.g., ECI; Tatsuoaka & Linn, 1983). These IRT-based indices usually provide the probabilities of correct responses and have their standardized versions. The group-based indices might include the Caution Index (CI) family (Sato, 1975; MCI, Harnisch & Linn, 1981; BW, Huang, 2011, 2012), the Dependability Index (Kane & Brennan, 1980), the U3 and ZU3 indices (van der Flier, 1982), and the Norm-Conformity Index (NCI, Tatsuoaka & Tatsuoaka, 1982). Most group-based indices are Guttman normed so that any response violating the Guttman model is deemed as aberrant. Having involved observed responses into the index definitions of aberrances, the CI family thus possesses an advantageous feature of corresponding to real response patterns, that is, index values and observed response patterns can be visually matched, not just overall probabilities provided by the IRT-based indices. Due to the development sequence, we will introduce the source of the CI family, the Sato caution index first and then the BW family indices, and how the BW family indices could be utilized in test construction for teachers in practice.

### **Sato caution index**

Sato (1975) developed the CI to measure the agreement of difficulty within a person/an item response pattern. It can be denoted by a complement of the ratio of the covariance between a vector that contains the observed binary item scores for an individual person and the ITEM-total vector to that between the Guttman vector of a person with a correct score number and the ITEM-total vector (Meijer, Muijtiens, & van der Vleuten, 1996), that is,  $CI_i = 1 - \sigma(X, n) / \sigma(X^*, n)$ . Since the CI taking the total score and item difficulty into consideration might detect both spuriously high and low response patterns in a single index, a washout effect of two different response patterns will occur. For example, in the analysis of a response pattern like (000|00000111), with "0" representing a wrong response and "1" representing a correct response, the within-ability response pattern (000 before the symbol "|") that might be caused by the psychological characteristic of "carelessness" will be deemed by the CI as identical to the response pattern beyond the ability level (111 after the symbol "|") in which the psychological characteristic of "guessing" might occur. That is, the two response patterns cannot be distinguished by the CI. Even though Harnisch and Linn (1981) had revised it to a modified CI (MCI) with the bound of 0 and 1, they still suffered from the ineffectiveness of differentiating extreme ability levels of students.

### **BW family indices**

The BW indices (Huang, 2011, 2012) inherited from the CI can effectively differentiate students (and items) with different levels of response aberrances by simultaneously providing misfit information on individuals and items for teacher-made tests used in classroom settings. Based on the Guttman assumptions (1944) in which a response pattern is deemed as ideal if responses are all correct within one's ability level (or all wrong within an item's difficulty level) and all wrong beyond one's ability level (or all correct beyond an item's difficulty level), the BW family indices expanded the CI measure to four types of indices ( $X_v^u$ ) in both person and item facets. In the BW family indices,  $X$  indicates the within- or beyond-index with a capital letter in person ( $v = i$ ) and a lowercase letter for item ( $v = j$ ),  $u$  refers to 1 for a correct response and 0 for a wrong response. Note that indices in item facet are just the symmetrical transformations of those in the person facet. Based on within and beyond levels of ability or difficulty, responses in

both facets can be attributed as normal or aberrant according to the levels of violation against the Guttman assumptions. The four types of indices ( $X_v^u$ ) in both facets are defined in Table 1.

In the person facet, the  $q_{.j}$  represents the conquering rates (the number of “0” in an item to the number of persons) ranked from low to high. The  $q_{iT}^*$  represents the average of conquering rates of the ( $T^{th}$ ) item and the ( $T+1$ )<sup>th</sup> item for the  $i^{th}$  examinee who earned a total score T in the number of items K. The bracketed expression, [(K-1)/2], is equal to the Gauss integer with the smallest integer greater than or equal to the value of (K-1)/2. Symmetrically in the item facet, the  $t_{ij}$  is the correct proportions of responses for each person (i.e., person ability) ranked in order from less able to more able (i.e., from bottom to top in the student problem S-P chart). The  $t_{Qj}^*$  represents the average person ability of the ( $Q^{th}$ ) person and the ( $Q+1$ )<sup>th</sup> person on the  $j^{th}$  item with difficulty level corresponding to a total item score Q in a sample of N examinees.

Table 1. BW family indices and interpretations

Response\Feature\ Facet		Person (ability level)		Item (difficulty level)	
Normal	Within	$W_i^I = \frac{\sum_{j=1}^{T_i} u_{ij} \times (q_{iT}^* - q_{.j})}{[(K-1)/2]} \times 100$	capability	$w_j^0 = \frac{\sum_{i=1}^{Q_j} (1 - u_{ij}) \times (t_{Qj}^* - t_{ij})}{[(N-1)/2]} \times 100$	difficulty
	Beyond	$B_i^0 = \frac{\sum_{j=T_i+1}^K (1 - u_{ij}) \times (q_{.j} - q_{iT}^*)}{[(K-1)/2]} \times 100$	misconception	$b_j^I = \frac{\sum_{i=Q_j+1}^N u_{ij} \times (t_{ij} - t_{Qj}^*)}{[(N-1)/2]} \times 100$	indistinctness
Aberrant	Within	$W_i^0 = \frac{\sum_{j=1}^{T_i} (1 - u_{ij}) \times (q_{iT}^* - q_{.j})}{[(K-1)/2]} \times 100$	carelessness	$w_j^I = \frac{\sum_{i=1}^{Q_j} u_{ij} \times (t_{Qj}^* - t_{ij})}{[(N-1)/2]} \times 100$	hint
	Beyond	$B_i^I = \frac{\sum_{j=T_i+1}^K u_{ij} \times (q_{.j} - q_{iT}^*)}{[(K-1)/2]} \times 100$	guessing	$b_j^0 = \frac{\sum_{i=Q_j+1}^N (1 - u_{ij}) \times (t_{ij} - t_{Qj}^*)}{[(N-1)/2]} \times 100$	Disturbance

In detail, two indices measuring “normal” responses within or beyond one’s ability level in the person facet are defined as the  $W_i^I$  index and the  $B_i^0$  index, respectively. They measure the extent of cognition mastering. The former is the *capability* index, named as such because it provides a measure of relatively conservative and authentic cognitive level for student mastery within his/her ability, whereas the latter is called the *misconception* index because it measures a student’s blind spots beyond the ability level in which he or she really misunderstood. However, two indices ( $W_i^0$  and  $B_i^I$ ) are designed to detect “aberrant” responses within or beyond one’s ability level for providing the degrees of concern based on easy items answered incorrectly or those surprising responses based upon difficult items on which the student answered them correctly; thus, they are called the *carelessness* index and *guessing* index, respectively.

Symmetrically, the item facet  $w_j^0$  index (*difficulty*) measures how well an item was structured as it relates to those test-takers whose abilities are really lower than the item’s difficulty level. The  $b_j^I$  index (*indistinctness*) measures how an item normally responded worse so that it cannot discriminate those test-takers whose abilities are much higher than its difficulty level. Finally, the two item-facet indices ( $w_j^I$  and  $b_j^0$ ) measure how an item aberrantly responded within or beyond its difficulty level, respectively. The  $w_j^I$  index refers to a “*hint*” measure because an item should conquer a person but did not. It, thus, might indicate the item provides some unrelated hint information of solutions so that students do not need the required capabilities but still can answer it correctly. However, the  $b_j^0$  index refers to a “*disturbance*” measure because an item luckily conquers qualified examinees with possible pitfalls.

The incorrect answers are not due to the item difficulty, but due to some deficient devices or obstacles underlying the item.

For person and item facets, the values of the normally responding indices ( $W_i^I$  and  $B_i^0/w_j^0$  and  $b_j^I$ ) range from 0 to 100; while those of the aberrantly responding indices ( $W_i^0$  and  $B_i^I/w_j^I$  and  $b_j^0$ ) range from 0 to 10. The larger the values, the more manifest the features.

Based on previous problem statements, this study, thus, aims at examining the quality of a teacher-made test on fractions and decimals and diagnosing students' related misconceptions through a classroom-based cognitive diagnostic model of the BW aberrance indices family. Three questions will be answered. (1) How well did the items of the teacher-made test on fractions and decimals function in terms of BW indices? (2) How well did the students master the fraction concepts? (3) How were the students classified on the basis of performances on these items?

## Method

### Design

The framework of this study was designed as Figure 1. From the teacher-made test, students' responses and teachers' item-concept data matrix were served as inputs for the analysis in the BW model. Based on person- and item- facet indices, the BW model can provide four kinds of feedbacks for teachers and students themselves, i.e., item analysis, students' response patterns, item mastery, and concept mastery.

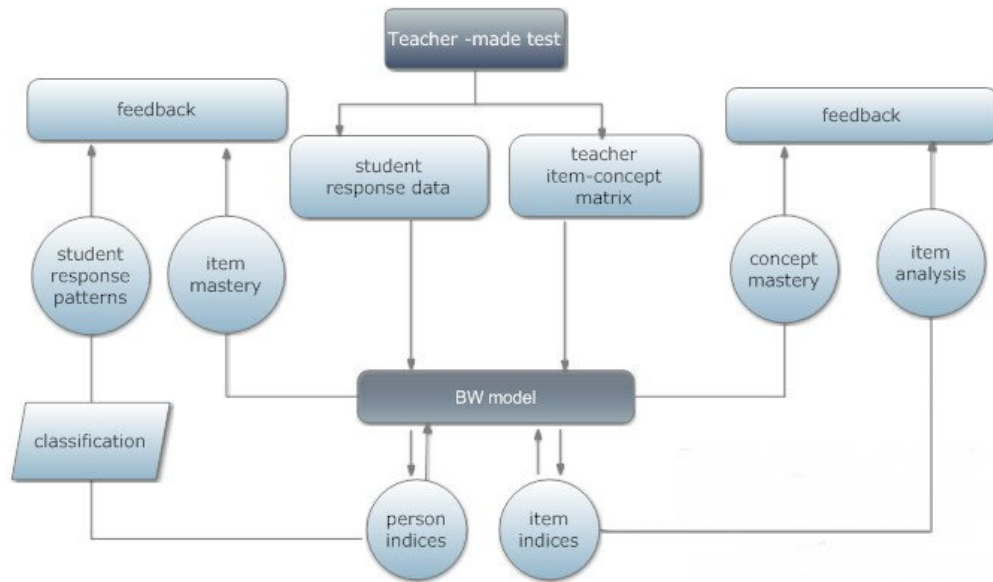


Figure 1. Research framework

### BW Model

Through integrating the indices in the person and item facets described previously, a probability model of answering item correctly can be postulated by the following three equations:

$$P_{ij} = \left\{ P_i \times (1 - P_j) \right\}^{u_{ij}} \times \left\{ (1 - P_i) \times P_j \right\}^{1-u_{ij}} \quad (1)$$

where

$$P_i = \left[ W_i^I \times (1 - W_i^0) \right] \times \left[ 1 - (1 - B_i^I) \times B_i^0 \right], \quad (2)$$



and

$$P_{.j} = [w_j^0 \times (1 - w_j^I)] \times [1 - (1 - b_j^0) \times b_j^I]. \quad (3)$$

Equation (2) indicates the overall ability level for the  $i^{\text{th}}$  examinee to answer correctly. Equation (3) indicates the overall difficulty level for the  $j^{\text{th}}$  item to conquer test-takers. Note that all the indices in person and item facets are calculated using the values without multiplying 100. Through the product of the person's ability probability with the compensation of the item difficulty probability with a 1/4 power on the product, the BW probability model is revealed in equation (1). Adding a 1/4 power on the product of probabilities in equation (1) is because the magnitude of each BW index is a proportion of all responses (values less than 1) and the product of indices in both equations (2) and (3) with a four-power expression of proportions will result in the product as singularly small.

The BW probability model of answering items correctly can be further extended to a cognitive diagnostic model. By estimating the BW probability model, we can obtain a probability matrix  $[R]_{ikj}$  that contains elements of the probability for persons answering items correctly. When multiplying the  $[R]_{ikj}$  matrix by the prerequisite item-concept matrix  $[F]_{jxk}$  that refers the attribution of the  $j^{\text{th}}$  item on the  $k^{\text{th}}$  concept, we can obtain a summative  $[C]_{ikx}$  matrix. Elements in the  $[C]_{ikx}$  matrix indicate the sum of probability that the  $i^{\text{th}}$  person mastering the  $k^{\text{th}}$  concept covered by the  $j$  items. Thus, dividing the summated probability by the total number of items in the test will yield an average probability for a person mastering a certain concept.

## Program

The WBstar program (Lu & Huang, 2010) was written in Visual Basic 6.0 to analyze cognitive oriented data. It includes item analysis and aberrance analysis. In the item analysis, WBstar can provide the values of the indices of item difficulty and discrimination and the distraction analysis as well as the KR20 reliability coefficient from the classical test theory (CTT). In the aberrance analysis, four parts of analyses can be provided: S-P chart from Sato, the values of BW family indices from item and person facets, BW item probability model, and BW concept probability model. The forms of data may be binary or raw to serve as response inputs to analyze the item analysis and aberrance analysis. For concept probability model estimation, it still needs an item-concept data input called the Q matrix (Tatsuoka & Tatsuoka, 1982) to identify the relationships of items to concepts. An example of an S-P analysis outcome is shown in Figure 2.

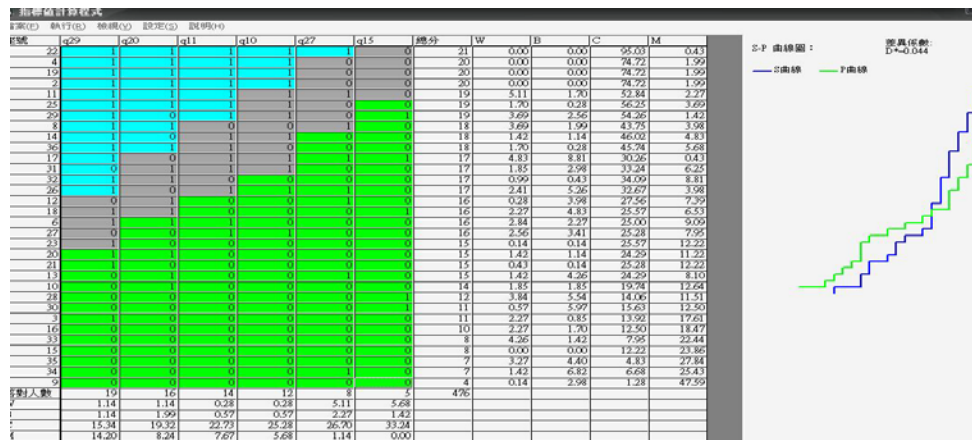


Figure 2. S-P analysis outcome

## Data

In this study, a teacher-made test on the unit of fractions and decimals was taken by 32 students in a fourth-grade elementary mathematics class in Taiwan. The test contained 22 items belonging to five fraction and decimal concepts: *transferring decimals into fractions (TDF)*, *equivalent fractions (EF)*, *comparing fractions with decimals (CFD)*, *transferring fractions into decimals (TFD)*, and *unit transformation (UT)*. Based on an agreed-upon judgment

proposed by two experienced teachers, a Q matrix of items on concepts (see Table 2) was set as follows and served in the computation of concept mastery probabilities.

Table 2. Q matrix of items on concepts

Item	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	q11	q12	q13	q14	q15	q16	q17	q18	q19	q20	q21	q22
Concept <i>TDF</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1
<i>EF</i>	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>CFD</i>	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1	1	0	0	0	0	0
<i>TFD</i>	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>UT</i>	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	1	0

## Classification

To provide more meaningful classification and interpretation, examinees were categorized into adequate groups based on the cutoff standards of the four person-facet indices. The approximate permutation test (Edgington, 1969) is an adequate approach to determine these cutoffs. The cutoff standards were calculated based on adequate percentages of index values from 100 permutations of the original  $32 \times 22$  data matrix. To avoid generating significantly dissimilar simulated matrices from the original data matrix, one can use a “weighted-permutation indicator (WPI) matrix” (Lu, Huang, & Fan, 2007). It is created by multiplying each original examinee’s total score to its fourth power with a randomly assigned probability and then by ranking the multiplication elements row by row across columns. Indexing this WPI matrix to the ranked original data matrix can form a re-ranked data matrix that is similar to the original one.

After generating similar simulated data matrices, three levels for the capability  $W_i^I$  index were set: H level (greater than 67% of permutations of indices), M level (between 33% and 66% of permutations of indices), and L level (less than 32% of permutations of indices). Two levels of the misconception  $B_i^O$  index were set based on the 50% of permutations of the indices. Regarding the carelessness  $W_i^O$  index and the guessing  $B_i^I$  index, the cutoff standards were set based on the 80% of permutations of the indices for the considerations of flagging alarms. Therefore, in total, there were 24 categories of examinees with different response patterns. Capital letters in categories represented the capability levels: high (H), average (M), and low (L); numbers represent types of response patterns: 1 for *Normal*, 2 for *Guessing*, 3 for *Careless*, and 4 for *Complex*. Categories represented typical response patterns, while those with an apostrophe (e.g., M1') represent quasi-typical response patterns and indicate that more concerns are needed for students’ learning blind spots or misunderstandings.

## Analysis

To evaluate the suitability of the test items, the disparity index ( $D^* = G/4NK\bar{p}(1-\bar{p})D_B(M)$ ) from the S-P chart analysis (Sato, 1982) is provided, where G is the total amount of “1” and “0” surrounded by the S curve and P curve; N is the number of students; K is the number of items,  $\bar{p}$  is the mean of correct ratios;  $D_B(M)$  is the constant array provided by Sato through the Gauss integer of the square root of NK plus 0.5. To examine how mastery students performed on fraction concepts, a one-way multivariate analysis of variance (MANOVA) was conducted. To examine how well the items performed and how well the persons responded, the item characteristic curves (ICC) and the person response curves (PRC) were plotted and discussed, respectively.

## Results

### Agreement examination

Through the S-P chart analysis, we can examine the homogeneity between student performances and problems’ functions. A value of .044 of the disparity index showed a good indication of agreement between the fourth-grade

students and this teacher-made test on fractions and decimals. That is, students' abilities were well assessed by the test, and these items were proposed to be suitable for these students.

### Item analysis

The CTT item analysis showed a good reliability quality for the teacher-made test on fractions and decimals ( $KR20 = .85$ ). Table 3 also provides item-facet aberrance indices ( $w_j^0$ ,  $b_j^I$ ,  $w_j^I$ ,  $b_j^0$ ), the index of item difficulty ( $P$ ), and the index of item discrimination ( $D$ ). Items in the fractions and decimals test displayed the levels of item difficulty from .13 to .94 and item discrimination from .00 to .88.

Table 3. Item-facet aberrance indices (ranked by  $W^0$ )

Item	$w_j^0$	$b_j^I$	$w_j^I$	$b_j^0$	$P$	$D$	Item	$w_j^0$	$b_j^I$	$w_j^I$	$b_j^0$	$P$	$D$
q1	0.00	84.94	0.43	0.71	.94	.13	q14	10.8	12.50	5.68	2.84	.69	.63
q2	0.00	84.94	0.43	0.71	.94	.13	q15	11.36	13.92	5.11	1.42	.63	.75
q18	0.00	85.23	0.43	0.43	.94	.13	q04	13.21	17.76	1.85	0.71	.56	.88
q08	0.85	72.44	0.00	0.00	.88	.25	q05	13.21	17.76	1.85	0.71	.56	.88
q13	1.14	67.61	0.14	0.71	.81	.38	q21	14.77	13.35	1.70	1.99	.50	.75
q3	1.14	59.38	0.57	4.83	.81	.13	q22	15.34	14.2	1.14	1.14	.56	.88
q09	2.56	53.69	0.57	2.84	.81	.38	q16	19.32	8.24	1.14	1.99	.44	.88
q10	5.11	11.65	11.36	3.69	.75	.25	q07	22.73	7.67	0.28	0.57	.44	.88
q17	5.26	33.95	2.13	4.12	.63	.50	q06	25.28	5.68	0.28	0.57	.44	.88
q19	9.38	25.00	1.70	3.13	.56	.63	q20	26.70	1.14	5.11	2.27	.25	.25
q12	9.66	26.14	1.42	1.99	.63	.75	q11	33.24	0.00	5.68	1.42	.13	.00

The order of the  $w_j^0$  index is reverse to that of the  $P$  index; that is, the indices measure similar properties of item difficulty, except in different directions. The increase of the  $w_j^0$  index means the increase of item difficulty, but the decrease of the  $P$  index. The property of the  $b_j^I$  index is reverse to that of the  $D$  index with desirable values revealed among adequate levels of item difficulty, but not good values among extreme ones (e.g., q1, q2, q18, and q11). Comparing to other acceptable items (e.g., q09 and q17), the q10 item, interestingly, with adequate levels of item difficulty ( $w_j^0 = 5.11$ ,  $P = .75$ ) should display desirable discriminations, but did not ( $b_j^I = 11.65$ ,  $D = .25$ ). The  $w_j^I$  index would further disclose possible underlying reasons for the q10 item:

[q10-q11]

John is  $160\frac{4}{5}$  centimeters tall; Jim is 1.605 meters tall.

q10: Who is taller?

The item q10 displayed the highest  $w_j^I$  value (11.36). Four students (see Figure 3, IDs of 34, 16, 28, and 30) responded to q10 with the answer of "John" not because they understood the question, but because of the values of numbers shown in it, that is,  $160\frac{4}{5}$  is greater than 1.605. This implied that item q10 provided unrelated information to the key concepts of solving this problem. Some students may not be familiar with the key concepts, but they can answer the item correctly just because of some obviously unrelated hint information revealed in the item. If the teacher changes the question so that the latter number is bigger, but with the same units, for example, 160.5 centimeters, the underlying hints may disappear.

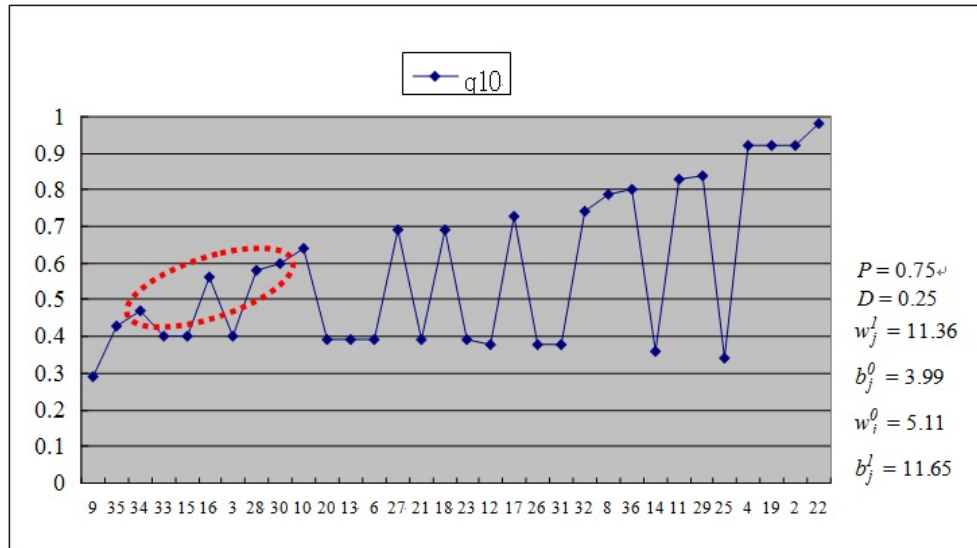


Figure 3. ICC for the q10 item

On the other hand, the item q3 displayed the highest  $b_j^0$  value (4.83, see Table 3). This implied that the information provided by item q3 might not be so sufficient or clear so that some students with high capabilities but answer it incorrectly. Items q1 to q3 are shown as follows.

[q1-q3]

A box contains 1000 balls. If Tom has 0.68 box of balls, then

$$0.68 = \frac{(q1)}{100} = \frac{(q2)}{1000}.$$

q3: Tom has ( ) balls.

Students easily answered items q1 and q2 correctly, but three of them (see Figure 4: q3 plot, ID8, ID3, ID28,) were confused by what q3 question really meant, especially for the capable ID8 student with the answer of 680/1000. Note that the answers came from or were related to q1 and q2. This indicated that the previous two items may confound the three students to solve the item q3. It is strongly recommended that item q3 be displayed solely in a test.

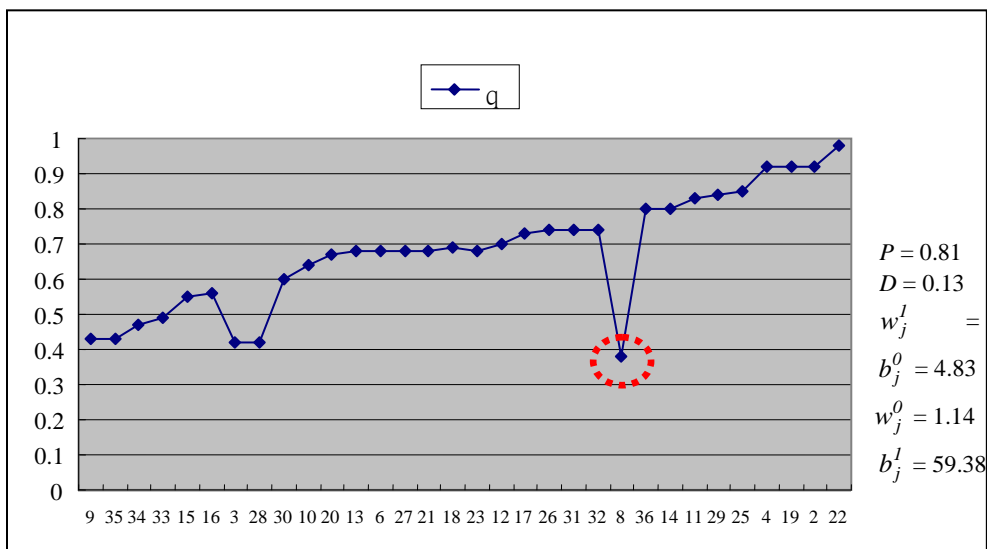


Figure 4. ICC for the q3 item

## Concepts mastery

There were no obvious interactions between ability level and mastering probability through the analysis of one-way MANOVA ( $Wilks' \lambda = .574$ ,  $F_{(8, 52)} = 2.079$ ,  $p = .055$ ). Thus, we examined the main effects of ability level and mastering probability, respectively. Although the sphericity assumption was violated, the within-subject effects still showed that the probabilities for students mastering the five fraction and decimal concepts were fairly different (*Lower-bound epsilon adjusted*  $F_{(1,29)} = 51.55$ ,  $p < .001$ , *partial eta square* = .64). In detail, most students mastered the concept of *equivalent fractions* (EF) with a mean of probability of .68 ( $SD = .17$ ), then the concept of *transforming decimals into fractions* (TDF:  $M = .62$ ,  $SD = .17$ ) or the concept of *comparing fractions with decimals* (CFD:  $M = .60$ ,  $SD = .19$ ), the concept of *unit transformation* (UT:  $M = .57$ ,  $SD = .18$ ), and least familiar with the concept of *transforming fractions into decimals* (TFD) with a mean of probability of .55 ( $SD = .17$ ).

For the between-subject effect, significant effects of ability were found ( $F_{(2,29)} = 98.376$ ,  $p < .001$ , *partial eta square* = .87). High-ability students mastered these concepts (mastering probability: .75–.86) greater than the middle level ones (.52–.70) and the low level ones (.35–.47) under a 95% confidence level. The strengths were consistently across the five fraction concepts (see Figure 5).

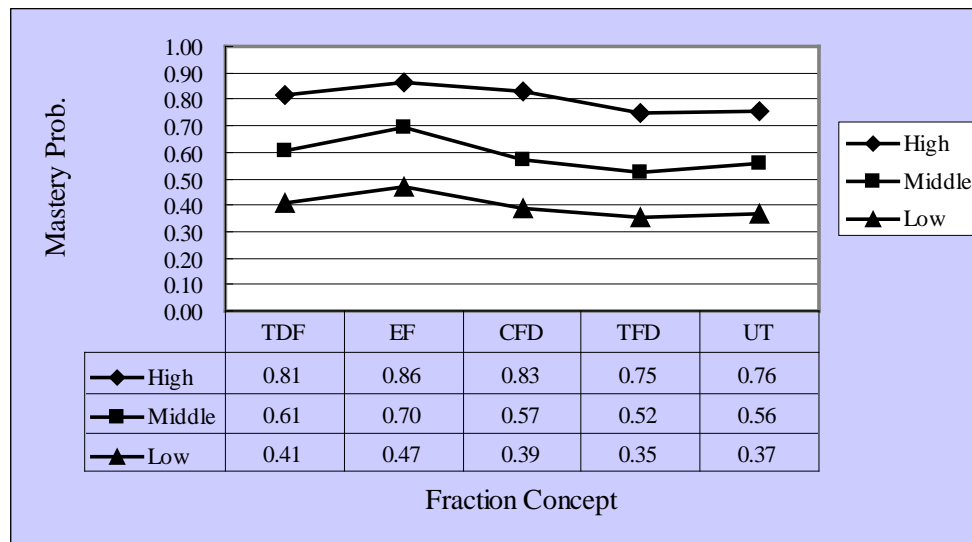


Figure 5. Mastery probabilities by student ability across fraction concepts

## Person responses classification

After the estimations of person-facet aberrant indices, 32 students were classified into 10 classes of response patterns (see Table 4). Among them, 21 students performed normally, four students tended to guess, four students tended to be careless, and three students exhibited complex response patterns, that is, combining both guessing and careless response patterns. Overall, the  $w_i^I$  index ranged from the least capable value of 1.28 to the most capable of 95.03 ( $M = 33.12$ ,  $SD = 23.11$ ). The misconceived  $B_i^O$  index ranged from 0.43 to 47.60 ( $M = 10.70$ ,  $SD = 10.11$ ). The value of the careless  $w_i^O$  index ranges from 0 to 5.11 with a mean of 1.83 and a standard deviation of 1.51. Finally, the value of the guessing  $B_i^I$  index ranges from 0 to 8.81 ( $M = 2.41$ ,  $SD = 2.33$ ). The levels of capability in different groups (H, M, and L) displayed a good distinction of range, that is, the  $w_i^I$  values greater than 43.75 belonged to the high-ability group, those between 19.74 and 34.09 belonged to the average-ability group, and those less than 15.63 belonged to the low-ability group. Students with the misconceptions  $B_i^I$  values greater than 8.81 (see M1' class) need to be concerned because their misunderstandings on these fraction and decimal concepts were relatively severe to their ability levels. In addition, students with the  $w_i^O$  and  $B_i^I$  values greater than 3.69 and 4.83 might tend to exhibit careless and guessing response patterns, respectively. Meanwhile, if those students with the values of the  $w_i^O$  and  $B_i^I$  indices were greater than 3.27 and 4.40, simultaneously, aberrances were flagged by complex responses for

combining both guessing and careless patterns. In addition, the values in Table 4 display corresponding characteristics of classes. For example, students classified in the M2 class tended to perform guessing responding behaviors and this scenario is satisfactorily reflected by the higher values of the minimum  $B_i^I$  index (4.83) comparing it to the normally responded class M1 (2.98) in the same ability level.

Table 4. Person-facet indices by class and by pattern

Index	$W_i^I$					$B_i^0$				$W_i^0$				$B_i^I$			
	N	Min	Max	M	SD	Min	Max	M	SD	Min	Max	M	SD	Min	Max	M	SD
Class																	
H1	7	45.74	<b>95.03</b>	66.74	18.13	<b>0.43</b>	5.68	2.94	1.86	<b>0.00</b>	1.70	0.68	0.86	<b>0.00</b>	1.14	.24	0.42
H3	3	<u>43.75</u>	54.26	50.28	5.70	1.42	3.98	2.56	1.30	3.69	<b>5.11</b>	4.16	0.82	1.70	2.56	2.08	0.44
M1	4	24.29	33.24	27.59	4.01	6.25	8.10	7.42	0.83	0.28	2.56	1.53	0.96	2.98	4.26	3.66	0.57
M1'	6	<u>19.74</u>	<u>34.09</u>	25.66	4.66	<u>8.81</u>	12.60	11.02	1.67	0.14	2.84	1.28	0.99	0.14	2.27	1.00	0.91
M2	2	25.57	32.67	29.12	5.02	3.98	6.53	5.26	1.80	2.27	2.41	2.34	0.10	4.83	5.26	5.05	0.30
M4	1	30.26	30.26	--	--	0.43	0.43	--	--	4.83	4.83	--	--	8.81	<b>8.81</b>	--	--
L1	4	<b>1.28</b>	13.92	9.98	5.85	17.60	<b>47.60</b>	26.90	14.08	0.00	2.27	1.17	1.27	0.00	2.98	1.38	1.27
L2	2	6.68	<u>15.63</u>	11.16	6.33	12.50	25.40	18.95	9.12	0.57	1.42	1.00	0.60	5.97	6.82	6.40	0.60
L3	1	7.95	7.95	--	--	22.40	22.40	--	--	4.26	4.26	--	--	1.42	1.42	--	--
L4	2	4.83	14.06	9.45	6.53	11.50	27.80	19.65	11.53	3.27	3.84	3.56	0.40	4.40	5.54	4.97	0.81
Pattern																	
Normal	21	1.28	95.03	36.74	24.97	0.43	47.60	10.67	10.36	0.00	2.84	1.11	0.98	0.00	4.26	1.33	1.44
Guessing	4	6.68	32.67	20.14	11.37	3.98	25.40	12.11	9.56	0.57	2.41	1.67	0.85	<u>4.83</u>	6.82	5.72	0.87
Careless	4	7.95	54.26	39.70	21.67	1.42	22.40	7.52	9.98	<u>3.69</u>	5.11	4.19	0.67	1.42	2.56	1.92	0.49
Complex	3	4.83	30.26	16.38	12.87	0.43	27.80	13.24	13.77	<u>3.27</u>	4.83	3.98	0.79	<u>4.40</u>	8.81	6.25	2.29
Total	32	1.28	95.03	33.12	23.11	0.43	47.60	10.70	10.11	0.00	5.11	1.83	1.51	0.00	8.81	2.41	2.33

### Aberrance patterns

All students were classified into 10 adequate classes based on their response patterns. In the upper part of Figure 6, we can see that typical highly capable students (H1 class) answered all easy items (left part from item q9 with difficulty 0.81) correctly and most failed in difficulty items (right part from item q16 with difficulty 0.44). Comparing to these typically responding students, however, three highly capable students in the H3 class (IDs of 11, 29, and 8) with aberrantly careless patterns tended to miss easy items (e.g., q3, q17, and q19).

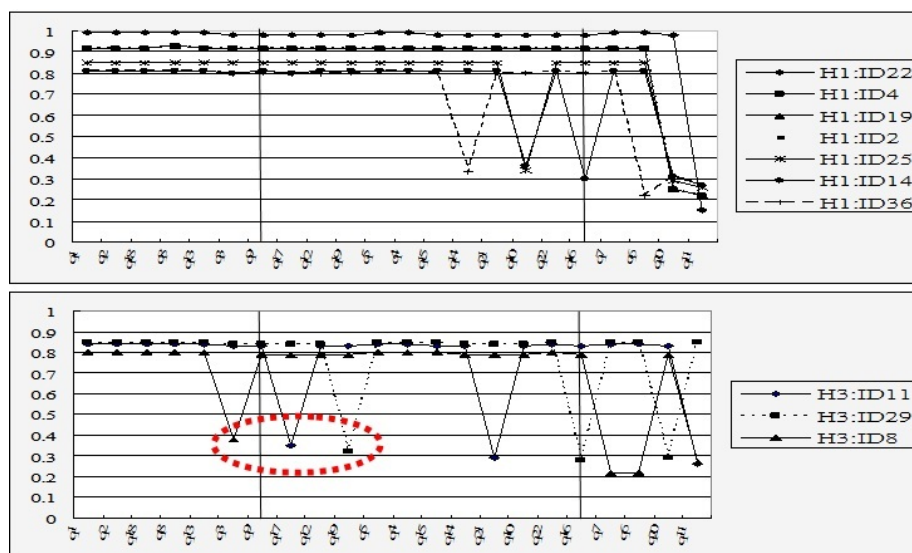


Figure 6. PRCs of High-level classes (H1, H3)

Next, the average capable students in the M1 class displayed typical response patterns in which most easy items were correctly answered and some medium to hard items were missed (see the upper part of Figure 7). A comparison of the M1 class response patterns to the M1' class revealed the characteristics of misconceptions, that is, almost all students in this class failed to understand the last four difficult items (the middle part). The students (IDs of 26 and 18) classified in the M2 class tended to guess items beyond their ability levels (the bottom part). The ID 17 student classified in the M4 class not only tended to guess hard items, but also tended to miss some easy questions (e.g., q12 and q19).

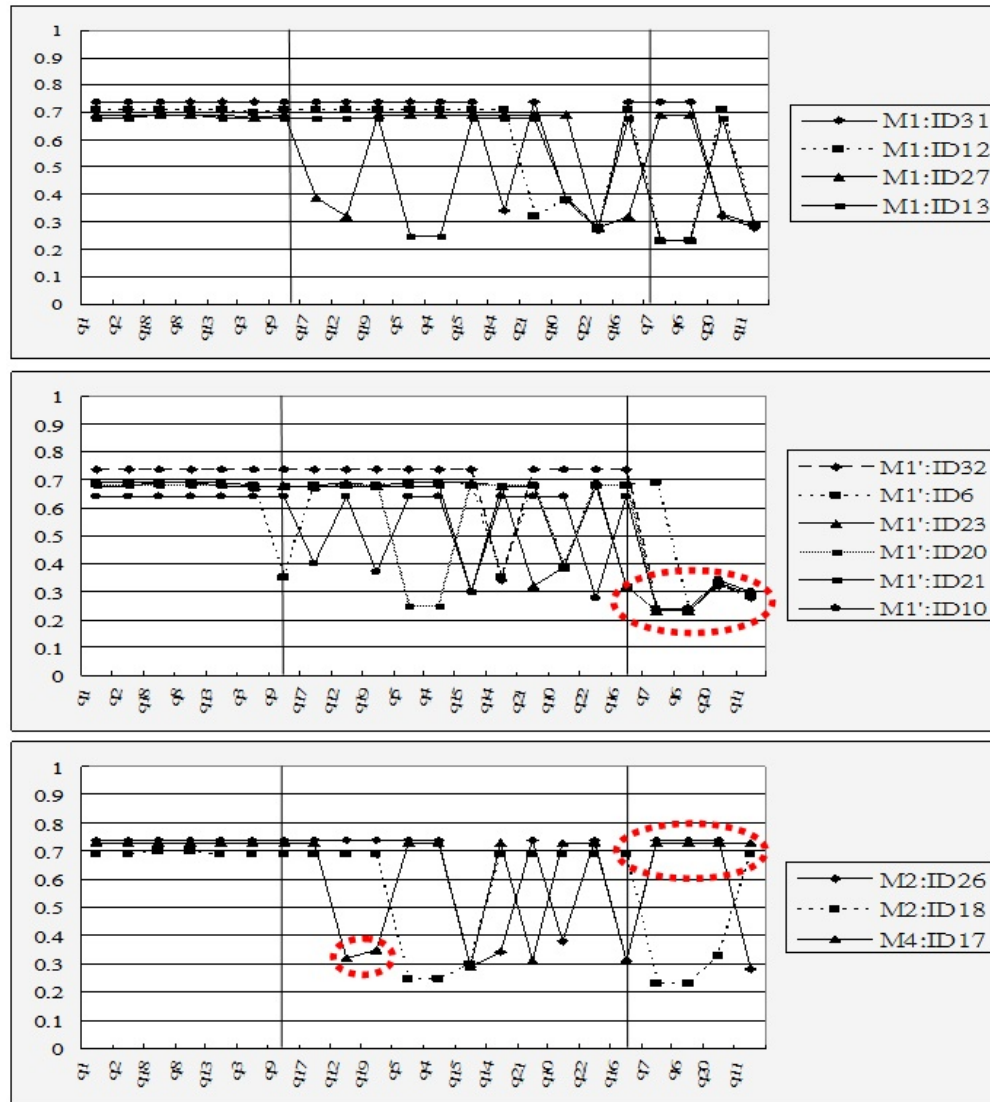


Figure 7. PRCs of Average-level classes (M1, M1', M2, M4)

Finally, the L1 class of students missed most medium to difficult items beyond item q9 (see Figure 8). The ID9 student was the least able student and even failed most easy items, especially in item q8 related to the transformation of decimals to mixed fractions. Students (ID30 and ID34) were grouped into the L2 class where they tended to guess items beyond their ability levels, while the ID33 student in the L3 class tended to miss the easiest items (see the middle part). Students (ID28 and ID35, see the bottom part) were classified into the L4 class because their responses contained both guessing and careless pattern.



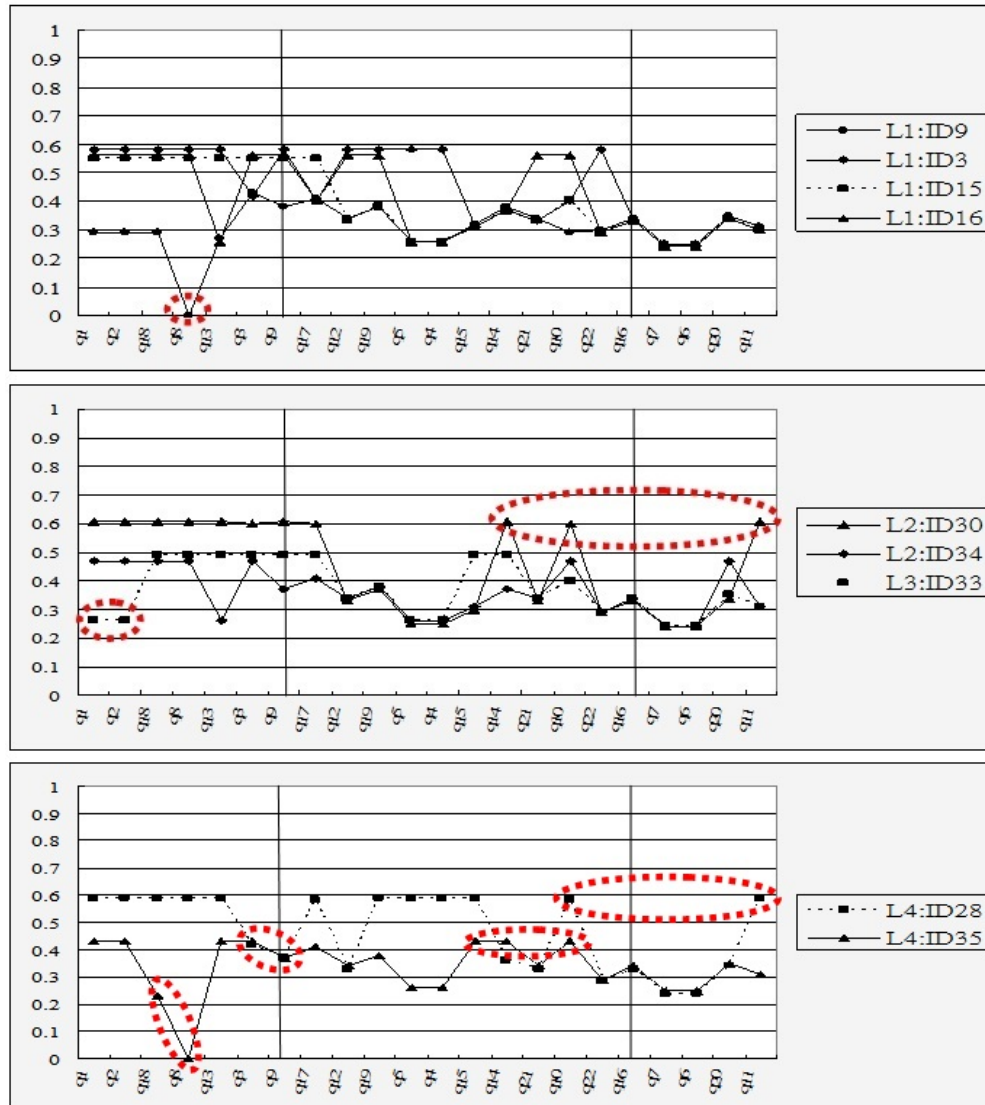


Figure 8. PRCs of Low-level classes (L1, L2, L3, L4)

## Conclusions

The study intended to apply the BW cognitive diagnostic model for teacher-made tests in the classroom. Unlike other CDMs that focus on the estimation of latent class parameters, the BW model was developed on the basis of interaction between a person's response properties (capability, misconception, guessing, carelessness) and an item's performance characteristics (difficulty, indistinctness, disturbance, hint). According to the taxonomy by Rupp and Templin (2008), the BW model might be categorized to both conjunctive and disjunctive condensation rules because of the combinations of indices. In both facets devices, the overall ability probability model and the overall difficulty probability model are based on the product of a conjunctive-condensation combination between the  $W_i^I$  index and the  $W_i^O$  index in the person facet (or  $w_j^O$  and  $w_j^I$  in item scenario) and a disjunctive-condensation combination between the indices of  $B_i^O$  and  $B_i^I$  in the person facet (or  $b_j^I$  and  $b_j^O$  in item scenario). Again, through the product of the person ability probability model with the compensation of the item difficulty probability model and with a 1/4 power on the product, the BW model can provide probability estimates of items correctly answered and concepts mastered.



In the person aberrance examinations, the  $W_i^I$  index and the misconception  $B_i^O$  index measured the extent of cognition to which a person was capable or misconceived in answering items, respectively. They were cognitive “content” oriented. However, the  $W_i^O$  index and the  $B_i^I$  index were cognitive “style” oriented so that a person tended to be careless or guess items, respectively. Symmetrically, in the item analysis, the hint  $w_j^I$  index measured the extent of obvious but unrelated messages to the key points of problem solving so that even an unable person could easily answer it correctly. The disturbance  $b_j^O$  index measured the extent of interferences or insufficiencies that an item exhibited so that a capable person failed it. The difficulty  $w_j^O$  index was similar to the item difficulty  $P$ , and the indistinctness  $b_j^I$  index measured the extent of inefficiency to which an item was not good enough to assess test-takers’ ability levels.

Based on the BW model, the study analyzed an empirical data matrix from a teacher-made test on fractions and decimals for a sample of fourth-grade students in Taiwan. All 32 students were classified into 10 classes and distributed in the normal, guessing, careless, and complex response patterns. With a good agreement between persons and items, the ICCs created by the probability model against persons’ capability levels showed effective examinations for the misfit of items. The PRCs generated by the probability model against items’ difficulty levels also showed effective examinations for persons’ response patterns. Through providing a Q matrix of items belonging to concepts, the probabilities of concept mastery were estimated and revealed the concept of *equivalent fractions* was most mastered and the concept of *transforming fractions into decimals* was least mastered.

The BW model can provide sufficient and matching information of interpretation to an original response data matrix in both person and item facets, especially in a small sample size situation like that of students taking a teacher-made test in class. Teachers can use the useful information provided by the WBstar program to improve their test development skills and to have immediate diagnostic feedbacks from their students’ response patterns. Moreover, the BW model can provide the extent to which students mastered concepts so that teachers can enhance their design of remedial instruction. However, some unknown features of the BW model and limitations still need to be disclosed. For example, the practical standards of item-facet indices are not as clear and simple as the index of difficulty  $P$  and the index of discrimination  $D$  interpreted. The classification rules for persons based on aberrance indices might be applied in the item scenario. In addition, the BW model in this study is only suitable for the dichotomously scoring situations; in future study, a generalized BW model can be developed for the use of data with polytomous scores.

## Acknowledgements

The authors would like to thank the National Science Council (NSC) in Taiwan for their financial support of this research.

## References

- Bezuk, N. S., & Bieck, M. (1993). Current research on rational numbers and common fractions: Summary and implications for teachers. In D. T. Owens (Ed.), *Research ideas for the classroom-Middle grades mathematics* (pp.118–136). New York, NY: Macmillan.
- Borasi, R. (1994). Capitalizing on Errors as “springboards for Inquiry”: A teaching experiment. *Journal for Research in Mathematics Education*, 25(2), 166-208.
- Dekker, T. (2007). A model for constructing higher-level classroom assessments. *Mathematics Teacher*, 101(1), 56-61.
- de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, 34(1), 115-130.
- Edgington, E. S. (1969). Approximate randomization tests. *Journal of Psychology*, 72(2), 143-149.
- Frey, B. B., Petersen, S., Edwards, L. M., Pedrotti, J. T., & Peyton, V. (2005). Item-writing rules: Collective wisdom. *Teaching and Teacher Education: An International Journal of Research and Studies*, 21(4), 357-364.

- Frey, B. B., & Schmitt, V. L. (2010). Teachers' Classroom Assessment Practices. *Middle Grades Research Journal*, 5(3), 107-117.
- Greer, B. (1988). Nonconservation of multiplication and division: Analysis of a symptom. *Journal of Mathematical Behavior*, 7, 281-298.
- Guttman, L. (1944). A basis for scaling qualitative data. *American Sociological Review*, 9(2), 139-150.
- Harnisch, D. L., & Linn, R. L. (1981). Analysis of item response patterns: Questionable test data and dissimilar curriculum practices. *Journal of Educational Measurement*, 18(3), 133-146.
- Hartman, H. J. (2001). Developing students' metacognitive knowledge and skills. In H. J. Hartman (Ed.). *Metacognition in learning and instruction: Theory, Research and Practice* (pp. 33-68). Dordrecht, The Netherlands: Kluwer.
- Hartnett, P., & Gelman, R. (1998). Early understandings of numbers: Paths or barriers to the construction of new understanding? *Learning and Instruction*, 8(4), 341-374.
- Holler, E. W., Gareis, C. R., Martin, J., Clouser, A., & Miller, S. (2008). Teacher made assessments: Getting them right. *Principal Leadership*, 9(1), 60-64.
- Hopkins, K. D. (1998). Educational and psychological assessment and evaluation (8th ed.). Needham Heights, MA: Allyn & Bacon.
- Huang, T. W. (2011). Robustness of BW aberrance indices against test length. *Knowledge Management & E-Learning: An International Journal*, 3 (3), 310-318.
- Huang, T. W. (2012). Aberrance detection powers of the BW and Person-Fit Indices. *Journal of Educational Technology & Society*, 15(1), 28-37.
- Kane, M. T., & Brennan, R. L. (1980). Agreement coefficients as indices of dependability for domain-setting judgments. *Applied Measurement in Education*, 1, 17-31.
- Kubiszyn, T., & Borrich, G. (2007). *Educational testing and measurement: Classroom application and practice* (8th ed.). Hoboken, NJ: John Wiley & Sons, Inc.
- Levine, M. V., & Rubin, D. B. (1979). Measuring the appropriateness of multiple-choice test scores. *Journal of Educational Statistics*, 4(4), 269-290.
- Linacre, J. M., & Wright, B. D. (1994). Chi-square fit statistics. *Rasch Measurement Transactions*, 8(2), 360-361.
- Liu, Y., Douglas, J. A., & Henson, R. A. (2009). Testing person fit in cognitive diagnosis. *Applied psychological measurement*, 33(8), 579-598.
- Liu, M. T., & Yu, P. T. (2011). Aberrant learning achievement detection based on person-fit statistics in personalized e-learning systems. *Educational technology & Society*, 14(1), 107-120.
- Lu, C. M., Huang, T. W., & Fan, J. R. (2007). Comparing the power of detecting aberrant response patterns by four Guttman based indices. *Psychological Testing*, 54, 147-174. (in Chinese)
- Lu, C. M., & Huang, T. W. (2010). *WBCM program* [Unpublished computer program].
- Mack, N. (1995). Confounding whole-number and fraction concepts when building on informal knowledge. *Journal for Research in Mathematics Education*, 26, 422-441.
- McGlohen, M., & Chang, H.-H. (2008). Combining computer adaptive testing technology with cognitively diagnostic assessment. *Behavior Research Methods*, 40(3), 808-821.
- Meijer, R. R., Muijtiens, A. M., & van der Vleuten, C.P. (1996). Nonparametric person-fit research: Some theoretical issues and an empirical example. *Applied Measurement in Education*, 9(1), 77-89.
- Meijer, R. R., & Sijtsma, K. (1999). *A review of methods for evaluating the fit of item score patterns on a test* (Research Report No. 99-01). Enschede, Netherlands: University of Twente, Faculty of Educational Science and Technology.
- Ni, Y. (2001). Semantic domains of rational numbers and the acquisition of fraction equivalence. *Contemporary Educational Psychology* 26(3), 400-417.
- Ni, Y., & Zhou, K. Y. (2005). Teaching and learning fraction and rational numbers: The origins and implications of whole number bias. *Educational Psychologist*, 40(1), 27-52.
- Nodoushan, M. A. (2011). The interface of interim assessment and feedback: An opinion. *Journal on Educational Psychology*, 4(3), 1-8.

- Rupp, A., & Templin, J. (2008). Unique characteristics of diagnostic models: A comprehensive review of the current state-of-the-art. *Measurement: Interdisciplinary Research and Perspectives*, 6(4), 219-262.
- Sato, T. (1975). *The construction and interpretation of S-P tables*. Toyko, Japan: Meiji Tosho. (In Japanese)
- Sato, T. (1982). *Practical uses of S-P Chart* [Elementary, Junior high, High school ed.]. Toyko: Meiji Tosho. (In Japanese)
- Seol, H. (1999). Detecting differential item functioning with five standardized item-fit indices in the Rasch model. *Journal of Outcome Measurement*, 3(3), 233-247.
- Smith, R. M. (1991). The distributional properties of Rasch item fit statistics. *Educational and Psychological Measurement*, 51(3), 541-565.
- Sowder, J. T. (1988). Mental computation and number comparison: Their roles in the development of number sense and computational estimation. In J. Hiebert & M. Behr (Eds.), *Number concepts and operations in the middle grades* (vol. 2, pp. 182-197). Reston, VA: National Council of Teachers of Mathematics.
- Tatsuoka, K. K., & Linn, R. L. (1983). Indices for detecting unusual patterns: Links between two general approaches and potential applications. *Applied Psychological Measurement*, 7(1), 81-96.
- Tatsuoka, K. K., & Tatsuoka, M. M. (1982). Detection of aberrant response patterns and their effect on dimensionality. *Journal of Educational Statistics*, 7(3), 215-231.
- van der Flier, H. (1982). Deviant response patterns and comparability of test scores. *Journal of Cross-Cultural Psychology*, 13(3), 267-298.
- William, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education Principles Policy and Practice*, 11(1), 49-65.

## Context-aware and Personalization Method in Ubiquitous Learning Log System

Mengmeng Li, Hiroaki Ogata\*, Bin Hou, Noriko Uosaki and Kousuke Mouri

The University of Tokushima, Japan // lemonrain99@gmail.com // ogaoga@mb2.tcn.ne.jp  
//myvstar@gmail.com // orchard.place@gmail.com // mourikousuke@gmail.com

\*Corresponding author

(Submitted January 20, 2012; Revised May 14, 2012; Accepted July 22, 2012)

### ABSTRACT

This paper explores a context-aware and personalization method in Mobile Learning system based on ubiquitous learning logs. Ubiquitous learning log stands for the log of knowledge or learning experience acquired ubiquitously. We construct a ubiquitous learning log system called SCROLL (System for Capturing and Reminding of Learning Log). Our research focuses on helping learners recall what they have learned making use of the learning contexts and learners' learning habits. Our measures consist of three main actions, which are to recommend learning objects in accordance with both learners' needs and contexts, to detect their learning habits using the context history and to prompt them to review what they have learned regarding their learning habits. What's more, by monitoring learners' reaction on the recommendation or prompting, the system can improve its prediction. An experiment was conducted to evaluate SCROLL and our method. The results of the experiment demonstrate that the learners benefited from the system and the context based recommendation and learning habit based prompting also motivated them to study more.

### Keywords

Personalized mobile learning, Ubiquitous learning log, Context-aware learning, Learning habit

### Introduction

Mobile technology has been believed holding out great promise for learning (Houser, Thornton, & Kluge, 2002). However, some of its limitations such as the small screen size, the high cost of 3G network and so on stopped the technology from growing as fast as we expected. In the last few years, a real great revolution is occurring in the mobile device world with the release of the new generation smartphones represented by iPhone launched by Apple Inc. and the open sourced Mobile OS Android released by Google. Since the new generation smartphones accommodate users with many advanced functions such as the multi-touch interface, full browser, GPS, millions of applications and so on, the number of smartphone users is increasing very sharply. Another key feature of smartphones is that they are equipped with a range of sensors such as the accelerometer, ambient light sensor, GPS, microphone, camera, compass and so on. Several years ago, researchers forecasted that the mass of mobile smartphones equipped with sensors could be turned into a giant distributed sensing system, allowing users to benefit from information gathered via other phones and users (Padmanabhan, 2008).

Our research primarily investigates the capabilities of the sensors of smartphones in context-aware and personalization mobile learning, because we find that the equipped sensors can play at least two important roles in our daily learning. On one hand, it can monitor learners' current context including their activities involving whether they are running, walking, listening to the music or surfing on the Internet and so on, and the environmental information including the location, time, temperature, and humidity and so on. Then, the system can recommend learning objects for a specified learner taking into account both the context and his study needs. On the other hand, it can track learners' contextual data as context history when they learn something using smartphones and catch individual's personal learning habits through analyzing the context history. Therefore, both learners' learning habits and their current context information will be interacted to support learning. These two basic ideas are applied to our method.

We developed a system called SCROLL (System for Capturing and Reminding of Learning Log) that allows learners to log their learning experiences with photos, audios, videos, location, QR-code (Quick Response code), RFID (Radio-frequency identification) tag, and sensor data and so on, and share and reuse them with others (Ogata Li, Hou, et al., 2010). The goals of SCROLL are lying in helping users to easily record their learning experiences and recall

them via the context, recommending other learners' learning experiences for them, finding out individuals' learning habits and supporting their learning in accordance with personal learning habits.

The rest of the paper is constructed as follows. Section 2 introduces our ubiquitous learning log system covering a primary scenario of its use, its main functions and architecture. In section 3, the workflow of the context-aware and personalization method is presented and then we explain it in terms of its three aspects. In section 4, we describe an experiment to evaluate the system and method, and the results of the experiment are given. Finally, we have a discussion about the results of the experiment and give the conclusions in section 5.

## **Ubiquitous learning log system**

With the evolution of the mobile devices, our lives are changing rapidly. For example, usually we take memos or notes (such as schedules, plans or task lists) in our pocketbooks. But now more and more people prefer to record these messages with their cell phones. For many people it is a simpler way, since the information can be stored in much more ways such as texts, photos, audios and videos and so on. Researches have also focused on facilitating this kind of "informal note taking" (Dai, Lutters, & Bower, 2005; Lin, Lutters, & Kim, 2004). However, besides informal note-taking we also take formal notes. For example, many language learners have vocabulary notebooks. We call these kinds of formal notes as learning logs. In this paper, a learning log is defined as a recorded form of knowledge or learning experience acquired in our daily lives and it serves as memory storage for notable or important knowledge to review, to remind and to reflect. In order to support such formal note taking and reminding, we design and implement our ubiquitous learning log system called SCROLL.

## **What is SCROLL**

The aim of SCROLL is to aid users to simply capture, review and reflect their learning logs, reuse the knowledge when in need, be reminded at right time at right place and be recommended others' learning logs properly. It adopts an approach of sharing user created contents among users. It means that a learner's learning log cannot merely be available for himself, but also can be shared with other learners who have the same learning needs. In our study, a learner's own and others' learning logs are his main learning objects.

To simplify the process of capturing the learning experience, the system provides a well-defined form to illustrate a learning log. It includes four basic elements, which are the time when the learning occurred (when), the knowledge (what), the sequence recorded in texts, photos, audios or videos that the learning should comply (how), and the location where the learning happened (where). Besides, the logs can be organized by tag and category. Figure 1(1) is the interface of adding a new learning log and Figure 1(2) is an example of learning log.

There is a special type of learning log called location based learning log in this study. This type of learning log is regarded as the knowledge that can be recalled by the location or place as a retrieval cue. Its purpose is to remind learners of what they have learned when they come to the place where the learning happened. According to the theory of encoding specificity, the place where we learned can be encoded as a retrieval cue initially and it is effective to activate a stored memory (Tulving & Thomson, 1973). For example, if we learned the Japanese names of vegetables in a supermarket, when we enter the supermarket next time some of what we have learned may come into our mind again.

SCROLL also provides some other functions such as navigator and Time map. Navigator is a function providing the learner with a live direct view of the physical real-world environment augmented by a real time contextual awareness of the surrounding learning logs (Figure 1(4)). Additionally, when a learning log is selected in the navigator, the system will show a path (route) for the learner to reach to the selected object from his/her current location (Figure 1(5)). Time map function means that the user can scroll the timeline above and then the map below will display the learning logs recorded during learners' selected period. It is designed to help the learners to reflect what they have learned. More detailed description on the functions of SCROLL system can be found in (Ogata, Li, Bin, et al., 2010).

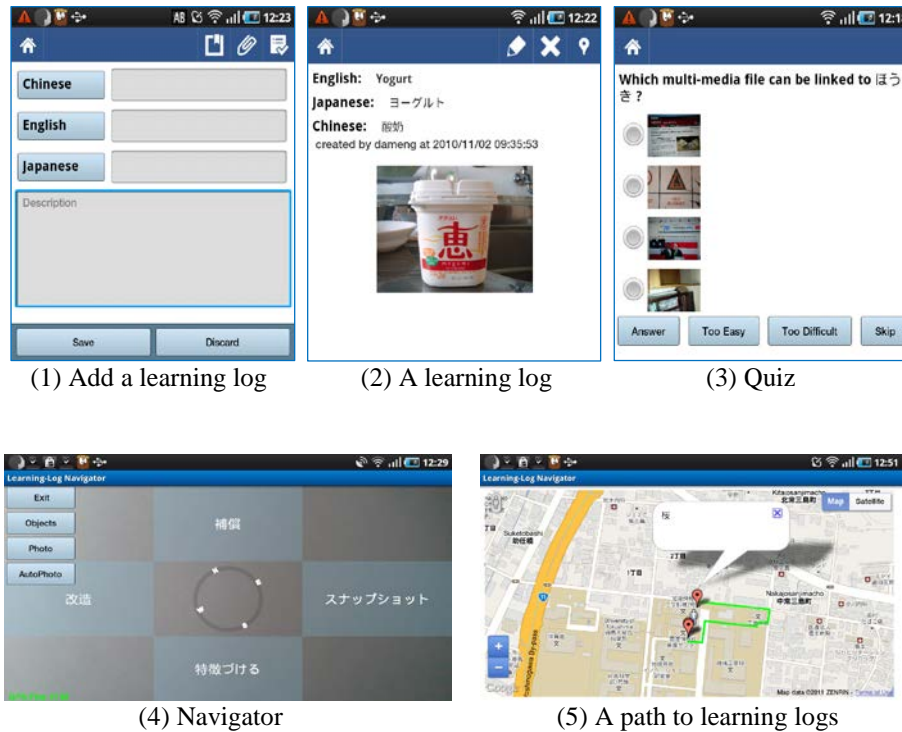


Figure 1. Interface of SCROLL

### The scenario of using learning log system

Up to now, SCROLL mainly focuses on language learning field. One typical scenario of its use is to assist international students to study Japanese in Japan. In this case, Japanese language learners, who face rich learning contexts every day, can gain much knowledge from their daily lives in different kinds of situations, such as shopping in the market, seeing a doctor in the hospital, having a haircut in a barbershop, visiting the museum and so on. They can not only take down what they have learned in those situations, but also will receive support from the system to recall and review them after that. Figure 2 illustrates the workflow of the scenario.

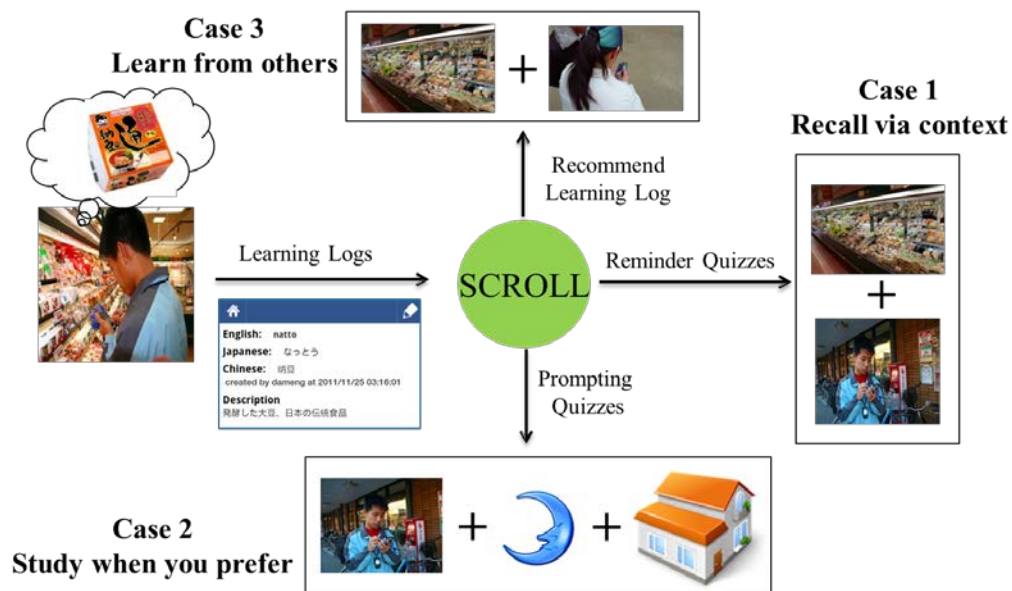


Figure 2. How SCROLL works

As shown, the learner learns a kind of traditional Japanese food called “natto” in a supermarket and he saves the knowledge as a learning log on SCROLL server. After that, there are three cases for the system to handle with the learning log:

1. Recall via context: When the learner enters the supermarket again, the system will provide him with reminder quizzes in order to help him recall “natto”.
2. Study when you prefer: If the system finds that the learner has a learning habit that he usually studies at home in the evening. And if the system detects that it is evening and the learner is at home, the system will prompt him to review what he learned.
3. Learn from others: If another Japanese learner enters the supermarket and she has the same language ability with the previous learner, the system will recommend the learning log about “natto” for her.

The context-aware and personalization method we propose in the rest of paper is responsible for these three cases.

### **Recall what we have learned**

An important goal of SCROLL system is to help learners recall what they have learned after they uploaded their learning logs. When a learner captures his learning log, besides the location based property mentioned above, a number of things are designed for learners to encode as retrieval cues. For instance, according to the picture superiority effect, the learning logs with pictures are much more likely to be remembered rather than those without pictures (Nelson, Reed, & Walling, 1976). In addition, according to the basic research on human learning and memory, practicing retrieval of information (by testing the information) has powerful effects on learning and long-term retention. And compared with repeated reading, repeated testing enhances learning more (Karpicke, Butler, Roediger, & others, 2009). For these two reasons, the quiz function taking advantages of the pictures, locations and so on is proposed. Three types of quizzes can be generated automatically by the system, which are yes/no quiz, text multiple-choice quiz and image multiple-choice quiz. Figure 1(3) shows an image multiple-choice quiz generated based on the meta-data of learning logs.

### **Context-aware and personalization method**

There are two objectives to adopt context-awareness and personalization in SCROLL:

- By being aware of a learner’s current context, especially the location information, the system can detect whether a learner is near to the place where he uploaded a learning log and whether there are location-based learning logs recorded by other learners close to him. If either requirement is met and the availability of the device is high, the system will show him a quiz based on the knowledge he gained around there or notify him the surrounding learning logs added by others.
- The system can record the context data when a learner uses the system to study as his context history and then catches his learning habits by making use of the context history. If the learning habits exist and the circumstance meets the learning habits, the system will show a piece of recommendation message to encourage him to review what he has learned.

To achieve these two goals, the system will monitor, analyze and dig learners’ contexts, derive the learning habits from them and prepare proper learning objects for them. Figure 3 demonstrates the whole processing flow of the method. It follows the below steps:

1. The system collects a learner’s context information from three parts: his activity, the status of device and the environmental data.
2. Then it analyzes the context and checks status of the device: for example, how much battery is left and whether the Internet is connected. If the availability is low, the system will do nothing.
3. If the device has a high availability, the system will check whether there is location-based knowledge near the learner. If existing, the system will provide location-dependent quizzes or recommend learning logs for him.
4. If there is no location-based knowledge for the learner, the method will examine if the learner is in his preferred learning context. If so, the system will show messages to encourage him/her to study.
5. All context data remains as context history to detect individual learners’ learning habits. Finally the learner’s response to the learning habits based recommendation is used to improve the learning habit detecting method.

The above processing flow reveals that the method consists of three terms, which are learners' current context, their learning habits and the learning objects. The following sections will introduce them respectively in detail.

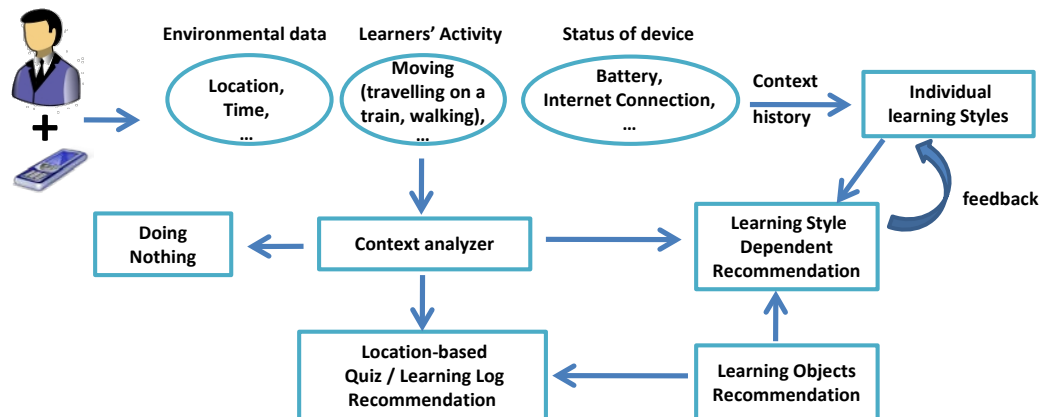


Figure 3. Workflow of the personalization and context-aware method

### Learners' context

A lot of studies on context-aware computing can be found in the literature. Here we'd like to highlight some of them. For example, Hwang and his team have done abundant of research in this field. Their studies cover learning activities based context-aware learning and learning tools supported context-aware learning. For example, the former type can be represented by learning in museum (Chiou, Tseng, Hwang, & Heller, 2010) and nature science observation activities (Hwang, Chu, Shih, Huang, & Tsai, 2010). The latter one includes Mindtools (Hwang, Chu, Lin, & Tsai, 2011), concept map (Hwang, Shi, & Chu, 2011; Hwang, Wu, & Ke, 2011) and an algorithm used for planning personalized learning paths (Hwang, Kuo, Yin, & Chuang, 2010) and so on. Ogata & Yano developed a system called JAMIOLAS which utilizes the environmental data to support Japanese learner to master the Japanese mimetic words and onomatopoeia (Ogata, Miyata, Hou, & Yano, 2010). Also some context-aware research explores location and time (or schedule) to recommend appropriate contents for learners (Clough, 2010; Nguyen, Hanoi, & Van Cong, 2012; Yau & Joy, 2010). By consulting these literatures, this study focuses the context on three aspects:

- Learner's activity: Learner's activity involves their motion (e.g. walking, running, travelling on the train or bus or sitting still) and what they do with the devices (e.g. listening to the music through earphone, surfing on the internet, or doing learning with SCROLL).
- Status of device: The status of device includes the battery, the Internet connection (3G, Wi-Fi or no connection), and the type of the ringtone (vibrate status or ringtone).
- Environmental data: The environment involves the location, time, temperature, weather and so on.

Based on the above data, the system will operate as follows:

1. Firstly, it will check the availability of the context. For example, whether the battery is enough (more than 20%), whether the Internet is connected, and whether the user is moving in a high speed and so on. 5km/h is thought as the average human walking speed (Knoblauch, Pietrucha, & Nitzburg, 1996) and exceeding 5km/h is treated as high speed.
2. If these conditions are satisfied, the system then will ascertain whether there are learning objects near him (within 50 meters) by using the location data. The learning objects include two parts: those he learned and those that he may want to learn.
3. If the learning objects he learned exist, the system will give a piece of message reading "Now you are near some learning logs you learned. Do you want to recall them in quizzes?" If he replied the message, the system will give him quizzes considering the place as a retrieval cue.
4. If the learning objects that meet his learning requirement, the system will show a piece of message saying "Some other learners' learning logs are found near here. Would you like to view them?" If they click the message, the system then shows a list of the learning logs and can navigate learners to the specified learning logs via navigator function.



5. Finally, the system will check whether the message is responded. If not responded, the system will recommend one more time when the user comes to the place again. But total the number of times of recommendation in the same area cannot exceed three times.

Additionally, the contextual data will be recorded as context history. In the literature, the context history, which is also believed to be useful, has not been fully utilized (Chen & Kotz, 2000). In our study, it will be reused for analyzing the learners' learning habits. We will introduce this part in the next section.

### Learners' personal learning habits

Learning habits are defined as a learner's habits when he learns. Learning habits play a very important role in learning because usually they are related to learners' daily customs and habits. With the mobile learning growing more and more important, many researchers have focused their studies on learners' learning habits or learning preferences. For example, Yau & Joy (2011) focus on learners' learning preferences including location of study, noise/distraction level in location and time of day. Hsieh & Lee (2012) employ learners' memory cycle, ability level and other learning preferences to support their English learning. Different from other researches, the learning habits supported in our system involve where a learner usually studies (such as home, school or fast-food restaurants), whether a learner has a habit of studying on the commuting train and when a learner prefers to study (e.g. after waking up in the morning or before sleeping at night). The context histories, collected when the learners use SCROLL, are utilized to detect whether a learner has any of the three learning habits. Three kinds of data including location, time, and speed are made use of. The following parts introduce the concrete method to detect the learning habits.

Since the time of learning every day is a discrete random value, we observed the regularity of the learning time in several periods to examine whether a learner has such learning habit or not. Concretely speaking, we separate a day into 24 phases. Each phase stands for an hour. Then we count the number of times of learning collected from a two weeks period in different phase. The next two periods of four weeks will be observed as well. Finally, the frequency phase which occupies more than 25% of the all learning times in three periods will be thought as the learner's preferred learning time. Figure 4 shows a participator's data in our experiment. It is obvious that the time from 23:00 to 24:00 is her favorite learning time.

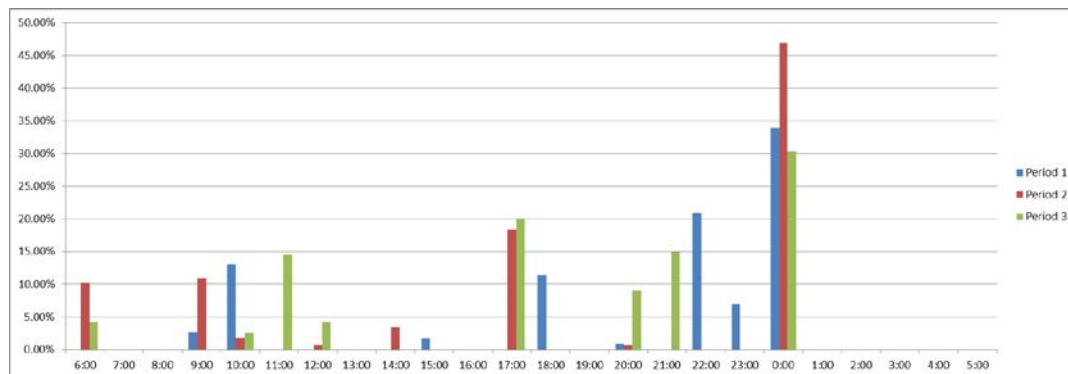


Figure 4. A learner's learning time in three phases

Considering how to find a learner's preferred learning place, many studies on finding users' significant place can be found (Ashbrook & Starner, 2003; Kang, Welbourne, Stewart, & Borriello, 2005; Zhou, Bhatnagar, Shekhar, & Terveen, 2007). We adopt a K-means algorithm. Firstly, we group a learner's learning locations into clusters. A cluster is a circle area whose radius is within 50 meters. Then, the scope that contains more than 30% of the learning location data is perceived as the learner's preferred learning location. Considering how to discover whether a learner has a habit of studying on a commuter train or bus, the speed and the time parameters are needed. Another experiential fact is that the speed and the time of commuting are relatively stable. Consequently, we firstly search the data with high speed (10~50 km/h is thought as the speed of the bus while above 50km/h is thought as the speed of the train (Toshiaki, Ryota, Hirokazu, & Tadashi, 2005)) and then group the data containing time and speed into clusters as well. The differences of the time is within 60 minutes and the difference of the speed is within 3km/h is

considered as the similar data. Therefore, if a cluster taking up 30% of the data existing, the system assumes this learner has the habit of studying on a commuter train or bus.

After achieved the learners' learning habits, the system can recommend messages when learners entered those environments. For example, when a learner stays in the place where he usually studies, a piece of message saying "The system guesses you are in a place where you usually do studies. Do you want to review what you have learned?" will be given. When it is his preferred learning time or when he is moving on a commuter train, he will receive a similar message as well. Finally, by checking the learners' response, the system can modify its prediction: if the system shows messages for him more than 3 times based on the same learning habit without any responses, this learning habit will be disabled.

## **Learning objects**

In this study, a learner's learning objects can be separated into two types: the learning logs that he learned (the ones that he uploaded or glanced through) and the ones recommended by the system. In order to provide learners with appropriate learning objects, when to remind them of the learned learning logs and what to recommend for them are two important issues. The former one is about the timing to show learners their learned learning logs in quizzes. The system adopts the graduated-interval recall method proposed by Pimsleur (Pimsleur, 1967). The intervals are 5 hour, 1 day, 5 days, 25 days, 4 months, and 2 years and so on. That is to say, after a learner added a learning log, the quiz about it will be available after 5 hours and then after 1 day and so on. Learners will be reminded continually. With respect to what to recommend, the system firstly takes into account the profile of the owner to recommend learning objects. It means that the learning log whose owner has the same study language and mother language will be recommended firstly. Also, the two learners should have same language abilities. And the recommended learning logs are followed in the order of reference times.

## **Evaluation and results**

### **Method**

There are two goals for us to evaluate the system and the method. Firstly we intend to investigate what kinds of benefits the system can bring for the learners and what kinds of improvements are needed by the learners. Secondly, we would like to observe how the learners act on the different kind of recommendation and what the limitations of the recommendation method are. To achieve the two goals, we conducted an experiment which lasted from April 15 to July 11 in 2011. 11 international students including 3 Chinese, 2 Taiwanese and 6 Korean participated. All of them have been in Japan for less than a year. The device adopted is Galaxy Tab SC-01C produced by Samsung. The transmission rate of 3G is 300 kb/s.

The experiment is divided into two steps. The first step (April 15~June 27) is used to evaluate the system. During this period, learners used the system to support their learning in daily lives. After more than 10 weeks of accumulating learners' learning history data, our experiment started the second step. This step is used to observe the context-aware and personalization method. It lasted two weeks. The first week was taken as a comparison, in which we sent each learner 3 pieces of recommendation messages randomly per day. The number of 3 is an approximate number on the basis of the presumption of the recommendation in the previous week. In the second week, the system provided them with the recommendation messages based on the analysis of the contexts and learning habits. After the experiment, the participators are asked to complete questionnaires.

### **Results**

This section we will introduce the results of our experiment by analyzing the user history data and the results of the tests and the questionnaires. The results of two steps of the experiment are introduced separately.

At the first step of the experiment, 1564 learning logs (AVE = 142.2, SD = 62.8) were uploaded and 4232 quizzes (AVE = 384.7, SD = 249.4) were done. That is, a learner records 1.95 learning logs and does 5.3 quizzes every day.

It means that they engaged in the system well. But high Standard Deviations reveal that each learner's involvement differs greatly. Hence, we decide to focus on statistics on every learner's usage of the system. In figure 5, x axis stands for the number of the memorized learning logs and y axis stands for the number of times of using SCROLL. If a learner answered correctly to the quiz about one learning log more than twice, this learning log is perceived as his memorized learning log. The number of times of usage consists of three parts: the number of learning logs that a learner saved, the number of times he did quizzes and the number of times of viewing learning logs. Figure 5 indicates that the more a learner engages in the system, the more learning logs he can remember.

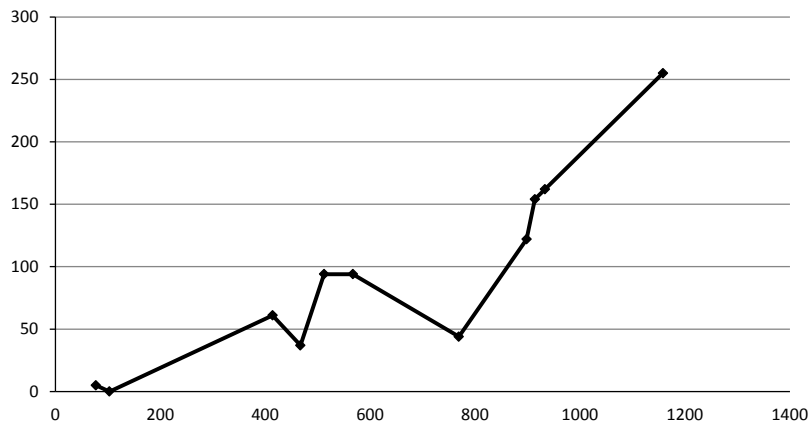


Figure 5. Number of memorized learning logs and number of times of using SCROLL

The questionnaire also asked the learners about how they evaluated the system. Table 1 shows the results (A five-point Likert-scale is used, the responses to which were coded as 1 = strongly disagree through to 5 = strongly agree.). From both Q1 and Q2, we can see that the system has a high usability and learners are satisfied with using it instead of paper.

Table 1. Questions about SCROLL in questionnaire

Number	Question	Score	SD
Q1	Is it easy for you to use SCROLL?	4.6	0.48
Q2	Is it better to use SCROLL than to use paper?	4	1.04

The second step of the experiment lasted two weeks. The first week is the control week and the second week is the experimental week. Table 2 presents an overview of the recommendation number and the response number in the two weeks. In the control week, the system randomly sent learners 224 pieces of recommendation messages and the participants responded only 19 pieces of them, rated 8.5%. In the experimental week, system sent 169 pieces of messages and the response rate increased to 33.1%. This increment demonstrates that comparing with the random recommendation, learners really benefited from the recommendation based on learning contexts and learning habits. Furthermore, according to the result of the questionnaire, 81.2% of the participants think the recommendation messages have stimulated them to learn more.

Table 2. The recommendation messages and the response

	Total number	Response number	Response rate
1 <sup>st</sup> week	224	19	8.5%
2 <sup>nd</sup> week	169	56	33.1%

As in the second week the recommendation method includes these three types, we would like to explore how the learners act on each type. Table 3 shows the total number of the recommendation messages, number of the response, response rate, average of evaluation score and the SD (Standard Deviation) of the evaluation score for three types of recommendation. From Table 3, we can see that learners responded the location-based quiz reminder messages most frequently and they also scored this type of recommendation highest.

Table 3. The recommendation messages and the response

Recommendation Type	Total	Responses	Percentage	Evaluation	SD
Location-based quiz reminder	87	35	40.2%	4.1	0.539
Location-based learning logs recommendation	72	18	25.0%	3.9	1.136
Learning habits based prompting	10	3	30.0%	4.0	0.632

### Comments from the participants

After the experiment, we also asked the participants to give their comments and share the episodes of using SCROLL system. Here we firstly pick out some positive comments and episodes:

- It is very helpful for me to memorize what I have learned by doing quizzes. Especially, the location based learning logs are very impressive.
- It is very interesting to be recommended other learners' vocabulary about food.
- It is very convenient to record what I have learned, especially when I travel. Also, it is easy to review what I have learned.
- The advantage of the SCROLL system for me is that it gave me the motivation to study Japanese. And I can easily retain what I have learned, because the pictures I took usually help me to recall them.
- As I had to take a JLPT (Japanese Language Proficiency Test), it helped a lot to prepare for the examination.
- Honestly every Sunday I wanted to relax the whole day. But when the recommendation came to me, I realized that I had to study Japanese. It kept my motivation.
- I always forget how to read a kind of noodle called "kamatama udon". But it is very helpful to review it before I order it in an "udon shop".
- When I talk in Japanese, sometimes I forget the word I learned. But with this system, it is easier to recall them.

From the positive comments and the episodes, we learn that learners benefit from the SCROLL system and the method well. The advantages such as easily capturing learning logs, reviewing old knowledge when in need, recalling via pictures and quizzes and keeping motivations via system recommendation and so on are mentioned in their comments. However, there are still some deficiencies reported by the participants:

- SCROLL system is very good. But I prefer to use paper, because it is more convenient.
- The battery is draining too fast. Especially, the GPS consumes too much power.
- The speed of Internet is too slow.
- The Galaxy Tab is a little big. It is inconvenient to bring it with me.
- As a Chinese, it is very easy for me to know the Kanji meanings but difficult to learn its pronunciation. However, the quizzes contain too many Kanji without its phonetic information. It will be more useful if it provides them.
- Sometimes the system alarms at classes. This embarrassed me.

From the above comments, we learn that battery, the Internet and the size of the device are unsatisfactory and the recommendation system is still with some limitations. In addition, some special functions are requested and some of them still prefer to use the traditional way to keep record of knowledge.

### Discussion and conclusions

With the results of the experiment and the comments provided by the participants, we would like to have a discussion about the study.

- According to the questionnaire, SCROLL is helpful for most of learners to record and review their learning logs. Compared with paper-based note taking, it facilitates learners retrieving their learning logs anytime and anywhere. However, there are still learners who prefer to use paper. This can be accounted for by the reason that some learners are more used to paper. For example, with papers they don't need to worry about the battery, internet speed and the usability of the system and so on. Besides, they can freely write on the paper, whatever they want to. On the contrary, in order to share with other learners and easily to be understood by the system, learners have to follow the template. This is a limitation of the system and it should be enhanced in the future work.

- The results of the experiment also prove that most learners benefit from the context-aware based recommendation and learning habits based prompting. The recommendation and prompting messages are provided in a relatively appropriate context and exert positive effects on keeping learners' motivation and recalling their old knowledge. However, the responses rates on the recommendation messages are relatively low and some users complained they were disturbed by the recommendation messages during classes. It is found out that it is difficult to catch concrete learners' contexts and their current learning motivation, because learners' willingness to study is always changing but the predication of the system is static. Consequently, it is necessary to empower the learners to control the recommendation method more so that they can customize the actions of the context-aware based and learning habit based recommendation. That way after the system achieved their learning habits, learners can view their own learning habits, and can change them. For example, a learner can determine the place where he usually studies in the map and he can change the place if he wishes. What's more, the learner can customize the action of the device when the system sends a recommendation message. For instance, he can set that the device will ring a tone or vibrate when the system shows recommendation messages.
- Through the experiment, we learned that in some cases it was difficult for a learner to master other learners' learning logs, even though they made correct answers to the quizzes. In our opinion, the learning log is a kind of knowledge with experience. Learners can learn more through experience than only viewing the other learners' experience. Therefore, it is necessary to help learners to experience other learners' learning logs. This is a very important issue in this study.

To conclude this paper, two aspects of things should be touched upon. Firstly, about the learning log system, we have found that most learners prefer to use this kind of tool for efficient learning, because it not only can store learners' learning contents but also help them to recall them. Besides, the usability of our system satisfied most of the learners. However, there are still some limitations such as the low speed of the 3G, the insufficient battery and the barrier to use the smartphones existed. But with the evolution of the mobile technology, this trend is inevitable. And besides language learning, other kinds of learning should be supported as well.

Secondly, learning log does not only contain knowledge and experience, but also record learners' learning context history information. Such information provides us a possible approach to know learners better, such as their learning habits. Besides, the smartphones can catch learners' current learning context. Our research has proved that these two kinds of data can be used to support learners' learning. However, because learners' activities are not always consistent with learning habits and the sensor technology cannot detect the learners' whole learning context perfectly, the learners' response rates to the recommendation are not satisfied. But we believe that the smartphones will be equipped more different kinds of sensors in the future and the device will know learners' better. As for our future work, we will try to use both learners' schedule and learning habits to catch learners' learning contexts more accurately.

## Acknowledgements

This research work was supported by JST PRESTO, and the Grant-in-Aid for Scientific Research No. 21650225 from the Ministry of Education, Science, Sports, and Culture in Japan.

## References

- Ashbrook, D., & Starner, T. (2003). Using GPS to learn significant locations and predict movement across multiple users. *Personal and Ubiquitous Computing*, 7(5), 275–286.
- Chen, G., & Kotz, D. (2000). *A survey of context-aware mobile computing research*. (Technical Report No. TR2000-381) Hanover, NH: Dartmouth College Publisher.
- Chiou, C. K., Tseng, J. C. R., Hwang, G. J., & Heller, S. (2010). An adaptive navigation support system for conducting context-aware ubiquitous learning in museums. *Computers & Education*, 55(2), 834–845.
- Clough, G. (2010). Geolearners: Location-based informal learning with mobile and social technologies. *Learning Technologies, IEEE Transactions on*, 3(1), 33–44.

- Dai, L., Lutters, W. G., & Bower, C. (2005). Why use memo for all?: restructuring mobile applications to support informal note taking. *CHI'05 extended abstracts on Human factors in computing systems*, 1320–1323. doi: 10.1145/1056808.1056906
- Houser, C., Thornton, P., & Kluge, D. (2002). Mobile learning: Cell phones and PDAs for education. *Proceedings of International Conference on Computers in Education* (pp. 1149–1150). doi: 10.1109/CIE.2002.1186176
- Hsieh, T. C., Wang, T. I., Su, C. Y., & Lee, M. C. (2012). A fuzzy logic-based personalized learning system for supporting adaptive English learning. *Educational Technology & Society*, 15(1), 273–288.
- Hwang, G. J., Chu, H. C., Lin, Y. S., & Tsai, C. C. (2011). A knowledge acquisition approach to developing Mindtools for organizing and sharing differentiating knowledge in a ubiquitous learning environment. *Computers & Education*, 57(1), 1368–1377.
- Hwang, G. J., Chu, H. C., Shih, J. L., Huang, S. H., & Tsai, C. C. (2010). A decision-tree-oriented guidance mechanism for conducting nature science observation activities in a context-aware ubiquitous learning environment. *Educational Technology & Society*, 13(2), 53–64.
- Hwang, G. J., Kuo, F. R., Yin, P. Y., & Chuang, K. H. (2010). A heuristic algorithm for planning personalized learning paths for context-aware ubiquitous learning. *Computers & Education*, 54(2), 404–415.
- Hwang, G. J., Shi, Y. R., & Chu, H. C. (2011). A concept map approach to developing collaborative Mindtools for context-aware ubiquitous learning. *British Journal of Educational Technology*, 42(5), 778–789.
- Hwang, G. J., Wu, P. H., & Ke, H. R. (2011). An interactive concept map approach to supporting mobile learning activities for natural science courses. *Computers & Education*, 57(4), 2272–2280.
- Kang, J. H., Welbourne, W., Stewart, B., & Borriello, G. (2005). Extracting places from traces of locations. *ACM SIGMOBILE Mobile Computing and Communications Review*, 9(3), 58–68.
- Karpicke, J. D., Butler, A. C., & Roediger, H. L. (2009). Metacognitive strategies in student learning: Do students practice retrieval when they study on their own. *Memory*, 17(4), 471–479.
- Knoblauch, R. L., Pietrucha, M. T., & Nitzburg, M. (1996). Field studies of pedestrian walking speed and start-up time. *Transportation Research Record: Journal of the Transportation Research Board*, 1538, 27–38.
- Lin, M., Lutters, W. G., & Kim, T. S. (2004). Understanding the micronote lifecycle: Improving mobile support for informal note taking. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 6, 687–694. doi: 10.1145/985692.985779
- Nelson, D. L., Reed, V. S., & Walling, J. R. (1976). Pictorial superiority effect. *Journal of Experimental Psychology: Human Learning and Memory*, 2(5), 523–528.
- Nguyen, V. A., Hanoi, V., & Van Cong, P. (2012). CAMLES: An adaptive mobile learning system to assist student in language learning. *Proceedings of Seventh International Conference on Wireless, Mobile and Ubiquitous Technology in Education* (pp. 72–76). doi: 10.1109/WMUTE.2012.19
- Ogata, H., Li, M., Bin, H., Uosaki, N., Moushir, M. E.-B., & Yano, Y. (2010). SCROLL: Supporting to share and reuse ubiquitous learning log in the context of language learning. *Research and Practice in Technology Enhanced Learning*, 6(2), 69–82.
- Ogata, H., Li, M., Hou, B., Uosaki, N., El-Bishouty, M. M., & Yano, Y. (2010). Ubiquitous Learning Log: What if we can log our ubiquitous learning? *Proceedings of the 18th International Conference on Computers in Education* (pp. 360–367). Retrieved from the ICCE website: <http://www.icce2010.upm.edu.my/contentpro.html>
- Ogata, H., Miyata, M., Hou, B., & Yano, Y. (2010). JAMIOLAS2: supporting Japanese mimetic words and onomatopoeia learning with wireless sensor networks for overseas students. *International Journal of Mobile Learning and Organisation*, 4(4), 333–345.
- Padmanabhan, V. (2008). Distributed sensing using mobile smartphones. *ITmagazine*, 22–24.
- Pimsleur, P. (1967). A memory schedule. *The Modern Language Journal*, 51(2), 73–75.
- Toshiaki, Z., Ryota, H., Hirokazu, A., & Tadashi, K. (2005, December). *Development of a method for estimating transportation modes with handy GPS equipment*. Paper presented at the 4th ITS Symposium, Tokyo, Japan.
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological review*, 80(5), 352–373.

Yau, J. Y. K., & Joy, M. S. (2010, March). *Designing and evaluating the mobile context-aware learning schedule framework: Challenges and lessons learnt*. Paper presented at the IADIS International Conference Mobile Learning, Porto, Portugal. Retrieved from <http://eprints.dcs.warwick.ac.uk/64/>

Yau, J. Y. K., & Joy, M. (2011). A context-aware personalised m-learning application based on m-learning preferences. *International Journal of Mobile Learning and Organisation*, 5(1), 1–14.

Zhou, C., Bhatnagar, N., Shekhar, S., & Terveen, L. (2007). Mining personally important places from GPS tracks. *Proceedings of International Conference on Data Engineering Workshop* (pp. 517–526). doi: 10.1109/ICDEW.2007.4401037

## Learning through Multi-touch Interfaces in Museum Exhibits: An Empirical Investigation

Panagiotis Zaharias<sup>1\*</sup>, Despina Michael<sup>2</sup> and Yiorgos Chrysanthou<sup>3</sup>

<sup>1</sup>Open University of Cyprus, Information Systems, Faculty of Pure and Applied Sciences, PO Box 12794, 2252, Latsia, Cyprus // <sup>2</sup>Cyprus University of Technology, Multimedia and Graphic Arts Department, 30 Archbishop Kyprianou Str. 3036, Limassol, Cyprus // <sup>3</sup>University of Cyprus, Computer Science Department, 75 Kallipoleos Street, P.O.Box.20537, CY-1678, Nicosia, Cyprus // pz@aueb.gr// despina.michael@cut.ac.cy // yiorgos@cs.ucy.ac.cy

\* Corresponding author

(Submitted January 23, 2012; Revised May 24, 2012; Accepted September 12, 2012)

### ABSTRACT

Interactive technologies are employed in museums to enhance the visitors' experience and help them learn in more authentic ways. Great amounts of time and money and many man-hours of hard work have been spent. But do such systems indeed achieve their goals? Do they contribute to a greater user experience (UX) and learning effectiveness? In this paper we describe the use of the "Walls of Nicosia" a 3D multi-touch table installed at the Leventis Municipal Museum in Nicosia, Cyprus. Two groups of students actively participated in this empirical study (they attended the 5th year class at elementary school, all aged from 10 to 11 years old): a) The traditional group (control group) where students took a guided tour throughout the museum and learned about the walls of Nicosia through printed maps exhibited at the museum and b) the virtual group where students interacted with the multi-touch application. The main aim of the study was to assess the learning performance and user experience between the two groups. Results showed no statistically significant differences in the learning performance but the virtual group reported user experience at significantly higher levels. The main findings are discussed and ideas for future research are presented.

### Keywords

Interactive learning environments, Evaluation methodologies, Elementary education, Virtual reality, Multimedia/hypermedia systems

### Introduction

Traditionally museums were using only real objects as exhibits to convey information to their visitors. During last two decades a new trend exists which is based on the involvement of the visitor and thus museums exploit technologically advanced systems to achieve their targets (Wishart and Triggs, 2010). The use of such systems aim in two directions: firstly to attract more visitors to the museums and secondly to "pass" the knowledge to their visitors in a more effective way.

Great amount in budgets and many man-hours of hard work and effort are spent to develop such systems. Does it really worth to do so? Do such systems indeed achieve their goals? Do they enhance the user experience (UX)? In this study we evaluate the learning performance and user experience of such a system installed at the Leventis Municipal Museum in Nicosia, Cyprus. The system under investigation is a 3D multi-touch table that runs an application about "the Walls of Nicosia" which is the fortification of Nicosia in different historical periods. We compare the results with those achieved with the traditional way of getting information from a museum which is by studying the exhibits (printed maps).

In the following section we describe related work on interactive systems installed at the museums. The Section 3 describes the multi-touch system that is evaluated and the application "The Walls of Nicosia" that runs on it. The next section, Section 4, is about the method that has been followed for the empirical evaluation. Results of the experiments are demonstrated at the Section 5. Finally at the last section we discuss our results and we give directions for further work on the subject.



## Related work

Technology is exploited recently by modern-day museums as a tool to convey information to their visitors; physical or virtual ones. Many museums setup technologically advanced systems in their physical space to attract more visitors and convey information in a more effective way or use online technologies in order to disseminate knowledge to remote visitors as well. Such systems are (or at least they should be) based on a theoretical background for learning and knowledge in the context of museum's content. In this section we give a brief theoretical background on learning related to technology and then we review work that has been done on interactive systems implemented in museums.

Hawkey (2004) made a review of aspects of learning provided by museums and galleries through the use of digital technologies. There are different learning philosophies regarding the learning opportunities in museums: should museums offer delivery or engagement? Should the underpinning rationale be a passive/transmission view or an active/constructivist view? Different taxonomies of learning experiences in a museum have been proposed. Gammon (2001) in his practical guide for museum evaluators classifies the learning process to cognitive, affective, social, skills development and personal categories. Hooper-Greenhill et al., (2003) proposed a quite similar set of learning experiences in a museum: a) knowledge and understanding, b) skills, c) values and attitudes, d) enjoyment, inspiration and creativity, and e) activity, behavior and progression.

During the recent years main emphasis has been put on a constructive process where the visitor feels that she is a privileged participant who has several possible learning pathways and possibilities. Accordingly, most of the proposed learning taxonomies take into account new learning developments and theories; such developments approach learning in the digital age as not a passive transmission of information to the learners/users of interactive technologies but as an active process through which people construct new understandings of the world around them (Resnick, 2001).

Such an approach can be verified by recent research findings that highlight the crucial role that technology and interactive systems can play in helping to maintain museums as constructive learning spaces. Interactive systems used by museums can be separated in two categories; those accessible from remote visitors, through online multimedia or VR systems and those systems that are located within the physical space of a museum. In the first category we have virtual museums, which extend the physical museums in a variety of ways (Bennet & Hodges, 2005). Focusing on 3D representations, a virtual museum may be a digital depiction of the existing museum and its exhibits (Kunkel & Averkiou & Chrysanthou, 2008; Patias, Chrysanthou, Sylaiou, Georgiades & Michael & Stylainidis, 2008), or it might include a combination of objects existing in several museums, while in some cases the 3D objects that have been modeled are based on description of historical documents (Gaitatzes & Christopoulos & Roussou, 2001). This category of systems, aims to convey information and knowledge that can be found at the museums, to virtual visitors who are not able to visit physically the museum.

Systems in the second category, that are located within the museum, aim to attract more visitors at the physical space of the museum. They provide modern ways of learning while increasing the satisfaction of museum's visitors. They allow the virtual interactivity of the user with a 3D representation of objects or the tour within virtual worlds. This category includes applications that run on VR systems (Roussou, 2001), Augmented Reality systems (Wojciechowski, Walczak & White & Cellary, 2004), haptics devices (Loscoc, Tecchia, Frisoli, Carozzino, Widenfeld & Swapp & Bergamasco, 2004), multi-touch tables (Averkiou & Chrysanthou, 2009; Geller, 2006) etc. These technologies combine in a great extent entertainment and education/ training.

Several studies have examined the aforementioned systems and applications for their learning strengths and affordances. Issues such as the learners' engagement and motivation, as well as the increase of learning performance and retention of knowledge have been investigated (Ang & Wang, 2006; Hut, 2007; Nijholt, 2000). More recently there is a focus on the provision of authentic learning experiences and the organization of learning activities. In the MuseumScouts project (Wishart & Triggs, 2010) a learner-centred approach in museums is adopted. In this project learners use information they collect during authentic learning opportunities in a museum to design short interactive multimedia teaching presentations with collaborative authoring tools.

Evaluation of such systems (Chittaro & Ieronutti & Ranons, 2004; Karoulis & Sylaiou & White 2006; Tzanavari, Vogiatzis, Zembylas & Retalis & Lalos, 2005) is another main research issue, since they are used by users with a great diversity in their profiles. Despite the increasing interest in the evaluation of museum interactive technologies, there is little knowledge regarding the crucial issue of user experience (UX) evaluation. Vavoula, Sharples, Rudman, & Meek & Lonsdale (2009) presented an evaluation of Myartspace, a service on mobile phones for inquiry-led learning. Such a service allowed students to gather information during a school field trip which is automatically sent to a website where they can view, share and present it, back in the classroom or at home. The evaluation during this study focused on usability issues, educational effectiveness and the impact of the new technology on school museum visits practice.

Reynolds & Walker & Speight (2010) described a three-stage qualitative evaluation programme of web-based museum trails in the Victoria and Albert Museum in London. The trails were only partially successful from a technological standpoint due to device and network problems. Nevertheless student feedback showed that overall the trails enhanced students' knowledge and their interest in the museums' objects.

### **The Walls of Nicosia**

The application "The Walls of Nicosia" is an interactive application that runs on a multi-touch table (Figure 1). The aim of this application is to allow the user to have a virtual tour through the fortifications of Nicosia across the centuries and learn about the Walls of the city at each historical period. The target was to explain and present the development of the area and the history of the development of the fortifications of the city – From a Roman and Byzantine Castle to a Medieval Royal Capital and a Venetian Fortified city.



*Figure 1.* The application "The Walls of Nicosia" runs on a multi-touch table.

The application uses 3D models (Figure 2) representing the fortifications of the city as well as the most important landmarks for five historical periods: pre-Roman (villages with no fortification), Roman castle, Byzantine castle, Lusignan Walls and Venetian Walls.



*Figure 2.* The fortifications of the city with the most important landmarks demonstrated with 3D models.

The interaction is achieved through the multi-touch surface of the table. A menu appears on the top of the screen of the table with a clickable button for each historical period. The users can navigate through time by touching using one finger on the corresponding button. They can then interact and study the 3D models of the historical period they choose using intuitive gestures, touching the surface with one or two fingers at a time. The users are able to zoom in/out, pan and tilt the virtual camera.

A helping menu exists on the interface of the application, describing which gesture should be used by the user in order to perform a specific operation. A compass, on the top-right corner of the interface, assists the users with orientation. Traditional music is played for each historical period, in order to immerse the user.

The multi-touch table hardware is based on back projection. A projector, connected with the host computer resides inside the table. The final image on the top surface is produced with the use of a mirror that reflects the projector image. This technique is used in order to virtually increase the space within the table and allow us to use a standard DLP project (not a short throw one). The gestures of the user's fingers are tracked using infrared light. 300 LEDs and an infrared camera have been mounted inside the table for the finger tracking. LEDs emit light toward the table surface. When a user touches the surface the infrared light is reflected back and the infrared camera captures that light. Each frame that is captured by the camera is processed in order to detect the position of the user's fingers on the screen.

### **Pedagogical affordances of “The Walls of Nicosia”**

The main purpose of “The Walls of Nicosia” is to provide rich interactivity and to facilitate users/visitors engagement and participation. “The Walls of Nicosia” was developed so as to provide several pathways to explore the walls of Nicosia during distinct historical times. Priority was given to provide an inquiry-based learning experience to the visitors. Inquiry learning is considered to be a very effective pedagogical strategy for a museum visit (McLeod and Kilpatrick, 2001). To this end, this interactive 3D application followed the pedagogical approach of guided-inquiry (Colburn, 2000). A typical scenario of guided inquiry in a museum prescribes a short introduction (by the teachers and museum guides) and then it is expected that students/visitors will be guided to uncover critical concepts for themselves and finally learn from the museum's objects rather than simply learn about them. “The Walls of Nicosia” provides opportunities for “hands-on” exploration while the 3D models of each historical period and the facility to zoom in/out, pan and tilt the virtual camera provides a greater sense of interactivity and authenticity to the users.

## **Methods**

### **Research design**

This study compared two different types of classes: (a) a traditional one (control group) where children took a guided tour throughout the museum and learned about the walls of Nicosia through printed maps exhibited at the museum and (b) the virtual group where children interacted with the multi-touch application “The Walls of Nicosia”. The main aim was to assess the learning performance and user experience between the two groups. The two different classes were based on the same learning content and learning objectives and both of them were located at the same place, the Leventis Municipal Museum of Nicosia. Accordingly in this study the following research questions were investigated:  
Are there any differences in the learning performance between students of the traditional group and the virtual group?  
Are there any differences in the user experience between students of the traditional group and the virtual group?

### **Participants**

The participants for the study were *randomly* selected from a list of elementary schools in Nicosia. In total 53 children (24 girls and 29 boys) participated in this study from 3 Elementary Schools of Nicosia. The children attended the 5th year class at elementary school, all aged from 10 to 11 years old.

## Data collection

Quantitative data were collected and observations were made by the authors that supported quantitative findings. Two questionnaires were developed:

- A pre-test questionnaire containing two parts, one part for collecting demographic information (age, gender, year at school) and the second part for evaluating learning performance, i.e., a knowledge test regarding the walls of Nicosia. The knowledge test contained 10 multiple-choice questions on issues related with the walls of Nicosia. This test was developed by the authors in collaboration with a colleague who is a school teacher and teaches topics on local history. The content validity was assessed by three experienced teachers who were responsible for the visit in the museum and they had very good knowledge about the history of Nicosia at different times.
- A post-test questionnaire containing two parts, one part for assessing the user experience and the second part for evaluating learning performance. As already mentioned the users in this field study were children. Therefore in order to assess the user experience we employed the Smileyometer and the Again-Again table (appendix). Both of them are tools contained in the FunToolkit (Read, 2008), which is a well-known and validated technique for assessing user experience with technology when the users are children (reliability analysis shown Chronbach's Alpha  $\alpha = 0.786$ , for the total scale and  $N = 53$ ). The knowledge test was the same as included in the pre-test.

Students' observation took place for both classes. The focus of this observation was to capture their motivation and experience as expressed by gestures, body movements, attention and words.

## Procedure

Concerning the procedure, the 53 students that participated in this field study were randomly (by using the lottery method) assigned to one of two groups: the traditional group (control group) and the virtual group. None of the students had ever been taught anything about the walls of Nicosia before. In addition as pointed out, both interventions (traditional class and virtual class) were based on the same learning content and learning objectives. Therefore we can assume that any differences in the learning performance and user experience can be attributed to the different type of classes (i.e. the two different interventions). Before starting the museum visit, students from both classes completed the questionnaire with demographic information (gender, age and year at school) and the knowledge test. Right after, the visit began in parallel for the two groups which last about 60-65 minutes. The visit experience included three parts: the first part was an introductory talk given by the teachers along with the presentation of the main exhibits (maps and multi touch table). The second part (20 minutes) was dedicated to a lively discussion and commentary on the Walls of Nicosia. The third part (20 minutes) focused on the assessment of User experience and the final knowledge test.

One researcher was present at each group. The traditional group entered the room with maps. Maps of the walls of Nicosia from different chronological periods are exhibited in this room. The teacher talked to the students about the fortifications showing to them the corresponding information on the maps (Figure 3).



*Figure 3. Exhibition of the printed maps at the traditional group*

The multi-touch table running the application was set up in a separate room. Students entered the room and a short introduction about the system followed by the researcher. Then the teacher talked to the students about the Walls. The teacher was interacting with the application while she was talking to the students and she was showing the corresponding information for the walls on the system.

Teaching and discussion in both groups was interactive. The teachers asked students several questions about the walls (e.g., their shape, extend etc) and prompted the students to express and share their opinions. The students were requested to find the answers by studying the maps for the traditional group or by interacting with the system for the virtual group. The researcher that observed the virtual group noted that there was a great enthusiasm and curiosity by the students. In cases where students did not give an answer, the correct answer was then given by the teacher. After the introduction and presentation of the system (first part of the visit), students in the virtual group were called to interact with the multi-touch table two at a time (second part). Afterwards students were given around 20 minutes (third part) in order to complete the post-test questionnaire by filling out the Smileyometer, the Again-Again table and the knowledge test.

## Data analysis and results

This study investigates a) whether there is a difference in learning performance between the students of the two different groups and b) whether there is a difference in user experience between the students of the two different groups. In order to identify any differences in the learning performance for the pre-test and the post-test phase an independent samples t-test analysis was performed. Significance level was set at 0.05 for all the analyses performed. For the investigation of changes in the learning performance within each group across the pre-test and the post-test phase, two paired-samples t-test were performed. Concerning the knowledge tests, a total score was calculated ranging from 0 to 10. Independent samples t-test was also employed in order to compare user experience between the two classes.

### Comparison of two groups regarding the learning performance

As far as concerns the pre-test phase, the t-test analysis did not revealed statistical significant differences [ $t(51) = -.304$ ,  $p = .762$ ] between the traditional group ( $M = 2.81$ ,  $SD = 1.642$ ) and the virtual group ( $M = 2.69$ ,  $SD = 1.258$ ) (Table 1). Such finding reveals that both groups had similar background knowledge on the walls of Nicosia.

Table 1. Pre-test learning performance

	Intervention	N	Mean	Std. Deviation	t	Sig. (2-tailed)
Pre-test knowledge score	Virtual group	26	2,69	1,258	-,304	,762
	Traditional group	27	2,81	1,642		

For the post-test phase the independent samples t-test analysis did not revealed statistical significant differences ( $t(51) = -1.889$ ,  $p = .065$ ) between the traditional group ( $M = 7.81$ ,  $SD = 1.902$ ) and the virtual group ( $M = 6.81$ ,  $SD = 1.980$ ). Although learning performance shows that traditional group performed higher at the post-test knowledge questionnaire (Table 2), this was a non statistical significant result.

Table 2. Post-test learning performance

	Intervention	N	Mean	Std. Deviation	T	Sig. (2-tailed)
Post-test knowledge score	Virtual group	26	6,81	1,980	-1,889	,065
	Traditional group	27	7,81	1,902		

Regarding the change in the learning performance within each group across the pre-test and the post-test phase, the two paired-samples t-test showed significant changes for both groups (Table 3). In the traditional group there was a

significant increase in learning performance from pre-test ( $M = 2.81$ ,  $SD = 1.642$ ) to post-test phase ( $M = 7.81$ ,  $SD = 1.902$ ),  $t(26) = -10.221$ ,  $p < 0.001$ ). Mean increase in learning performance was 5,000. In the virtual group there was also a significant increase in learning performance from pre-test ( $M = 2.69$ ,  $SD = 1.258$ ) to post-test phase ( $M = 6.81$ ,  $SD = 1.908$ ),  $t(25) = -9.474$ ,  $p < 0.001$ ). Mean increase in learning performance was 4.115 (Table 3), which was slightly smaller than the increase in the traditional group.

*Table 3. Changes in learning performance*

					Paired differences		
		Mean	N	Std. Deviation	Mean	T	Sig. (2-tailed)
Traditional group	Pre-test learning performance	2,81	27	1,642	-5,000	-10,221	,000
	Post-test learning performance	7,81	27	1,902			
					Paired differences		
		Mean	N	Std. Deviation	Mean	T	Sig. (2-tailed)
Virtual group	Pre-test learning performance	2,69	26	1,258	-4,115	-9,474	,000
	Post-test learning performance	6,81	26	1,980			

### Comparison of two groups regarding the user experience

We move on with the analysis of the second research question which is about the user experience. Independent samples t-test was also employed in order to compare user experience between the two classes. For the Smileyometer, the analysis showed statistical significant differences [ $t(51) = 3.042$ ,  $p = .004$ ] between the traditional group ( $M = 3.96$ ,  $SD = 0.898$ ) and the virtual group ( $M = 4.65$ ,  $SD = 0.745$ ). It is evident that students from both classes reported high levels of user experience, however findings from this analysis shows that virtual group reported user experience at significantly higher levels (Table 4). For the Again-Again table, analysis also revealed statistical significant differences [ $t(51) = 2.947$ ,  $p = .016$ ] between the traditional group ( $M = 2.33$ ,  $SD = 0.452$ ) and the virtual group ( $M = 2.73$ ,  $SD = 0.679$ ) (Table 4).

In order to control other variables for possible effects on the results, additional test were performed. In more details, we tested whether gender had a significant impact on learning performance and user experience. No statistical significant differences were found for learning performance as for the pre-test [ $t(51) = 1.516$ ,  $p = 0.136$ ] and the post-test [ $t(51) = -1.070$ ,  $p = 0.290$ ]. No statistical significant differences were found for user experience [ $t(51) = -0.693$ ,  $p = 0.492$ , for the smileyometer] and [ $t(51) = .144$ ,  $p = 0.886$ , for the again-again table].

*Table 4. User/visitor experience*

	Intervention	N	Mean	Std. Deviation	T	Sig. (2-tailed)
UserExperience (smileyometer)	Virtual group	26	4,65	,000	3,042	,004
	Traditional group	27	3,96	,833		
	Intervention	N	Mean	Std. Deviation	T	Sig. (2-tailed)
UserExperience (again-again table)	Virtual group	26	2,73	,452	2,947	,016
	Traditional group	27	2,33	,679		

### Observations

It was observed that students participating in the traditional group were passively listening to the teacher without any active role in the learning procedure. On the other hand students in the virtual group were enthusiastic and actively

engaged in the whole process. This finding is also confirmed by the quantitative analysis as far as concerns the user experience analysis.

The discussion in the virtual group was very lively and kept the interest of the students during the whole visit. For instance, as soon as the teacher of the virtual group switch to the Venetian period a student said *“oh now I can see the shape of the ramparts”* which was one of the questions of the knowledge test while another student started to count how many ramparts exist that was the issue of another question. Some other instances from the process: When the teacher pointed out on the display the shape and direction of the river we observed that the students came nearer to the table to be able to see by themselves the river and interact with it. In addition another student stated that *“the Walls are very big!”* and the teacher confirmed his observation saying that they indeed cover several square Kilometers.

## Discussion and future research

In this study, “The Walls of Nicosia”, a 3D multi-touch table which is part of the permanent exhibition at the Leventis Municipal Museum of Nicosia, is presented. The main focus is on investigating the user experience and the learning effectiveness after the interaction with this interactive technology. Two groups of students, who were visiting the museum, participated in the study. In the virtual group students interacted with the 3D multi-touch application in order to learn about the history of walls in old Nicosia, while in the traditional group students took a guided tour throughout the museum and learned about the walls of Nicosia through printed maps exhibited at the museum.

The results showed no statistically significant differences in the learning performance between the two groups. Although results of learning performance shows that traditional group performed higher, this was a non statistical significant result.

Initially it was expected that students in the virtual group would perform better; many other studies (Wishart & Triggs, 2010; Tüzün, Yilmaz-Soylu, Karakus & Inal & Kizilkaya, 2009; Ke & Grabowski, 2007) with students have showed that interaction with 3D applications, such as games, virtual worlds etc. can contribute to higher learning performance across several topics (history, geography, mathematics, literature). Although the learning performance was increased in a statistically significant way in both groups (comparing knowledge before and after the museum visit), there was no significant difference in the learning performance between the two groups. Other studies as well (Wrzesien & Raya, 2010; Papastegriou, 2009) demonstrated no differences in the learning performance, mainly due to the perceived novelty of the interactive technology at hand. For instance in this study, the innovative way of exploring and navigating the 3D multi touch interactive environment may explain the learning performance of the virtual group. The several novel features and the realistic 3D graphics of the “Walls of Nicosia” may have distracted to a certain extent the students in the virtual group. On the other hand the physical tour at the museum’s exhibition seems a more natural way of learning about the history of walls of the old Nicosia. This was more or less confirmed by the informal observations of the researchers and the commentaries made by the students after their visit.

As far as concerns the user experience issues, findings from this study shows that virtual group reported user experience at significantly higher levels (as depicted in Smileyometer instrument). They seem to enjoy it more than the other students, being more engaged and they expressed a greater intention to repeat such a visit (as drawn from the Again-Again table). Such findings confirm relevant data from other studies as well (Wrzesien and Raya, 2010; Papastegriou, 2009).

The study has some specific limitations. The knowledge test we developed was quite short and focused on factual type of knowledge and short-term retention. In a future study a longer-term retention of knowledge and other types of knowledge (other than simple concepts and mere facts) can be evaluated. To this end, future studies can perform some post-museum visit activities in order for the students to reflect upon what has been learned during their museum visit and further assess whether knowledge can be retained.

Regarding the organization of the field study: The fact that students in the virtual group were called to interact with the multi-touch table two at time may have influence the user experience. Additionally there were two different

teachers who were responsible for the museum visit for the respective groups of students. Despite the fact that the learning content and objectives were the same and the protocol for the “teaching” process, the presentation of the exhibits by two different persons could influence the results in both the learning performance and the user experience.

At the current state the “Walls of Nicosia” does not provide the opportunity to build an avatar, so as to enhance the perception of presence and the ownership of the virtual environment. Such enhancements can lead to a greater contribution to the learning effectiveness and the user/visitor experience; this could be investigated in a future study. Moreover, a future study can pursue the deeper investigation of the impact that individual museum’s visitor characteristics (such as specific learning and cognitive styles) may have on the learning process and the whole user/visitor experience. Moreover, a future work could include an assessment of anticipated user experience of both groups, which can further enlighten the focal research questions.

In conclusion, this study demonstrates that the use of new types of interactive systems contribute to the experience of visitors in museums, enhancing their level of active participation and engagement and their intention to repeat visits. As for the learning gains that the visitor should grasp, it can be argued that such interactive technologies provide new learning experiences no less than the traditional exhibition methods. In many cases and after the novelty effect has passed, such interactive technologies can provide more authentic learning and entertainment at the same time.

## Acknowledgments

We would like to thank all the students and their teachers for their comments and fruitful participation as well as the people in the Leventis Municipal Museum of Nicosia for their support (<http://www.leventismuseum.org.cy/>).

## References

- Ang, K., & Wang, Q. (2006). A case study of engaging primary school students in learning science by using active worlds. In R. Philip, A. Voerman, & J. Dalziel (Eds.), *Proceedings of the First International LAMS Conference 2006: Designing the Future of Learning* (pp. 5-14). Sydney, Australia: The LAMS Foundation.
- Averkiou M., & Chrysanthou Y. (2009). Evaluating a multitouch interface for 3d navigation inside the virtual world of a museum exhibit. In K. Debattista, C. Perlingieri, D. Pitzalis & S. Spina (ed.), *Proceedings of Virtual Reality, Archaeology and Intelligent Cultural Heritage*. Aire-la-Ville, Switzerland: The Eurographics Association.
- Bennet, J., & Hodges, A. (2005). The development of science museum web sites: Case studies. In L. T. W. Hin, & R. Subramaniam (Eds.), *E-Learning and Virtual Science Centers* (pp. 366-392). Hershey, PA: Information Science Publishing. doi:10.4018/978-1-59140-591-7.ch018
- Chittaro, L., Ieronutti, L., & Ranon, R. (2004). Navigating 3d virtual environments by following embodied agents: A proposal and its informal evaluation on a virtual museum application. *PsychNology Journal: Special issue on Human-Computer Interaction*, 2(1), 24-42.
- Colburn, A. (2000). An inquiry primer. *Science Scope*, 23(6), 42-44.
- Gaitatzes, A., Christopoulos, D., & Roussou, M. (2001). Reviving the past: cultural heritage meets virtual reality. In D. B. Arnold, A. Chalmers, & D. W. Fellner, *Proceedings of the 2001 conference on Virtual reality, archeology, and cultural heritage* (pp. 103-110). New York, NY: ACM
- Gammon, B. (2001). *Assessing learning in museum environments: A practical guide for museum evaluators*. London, UK: Science Museum.
- Geller, T. (2006). Interactive tabletop exhibits in museums and galleries. *IEEE Computer Graphics Applications*, 26(5), 6-11.
- Hawkey, R. (2004). *Learning with digital technologies in museums, science centres and galleries* [Report 9, Futurelab Series]. London, UK: King’s College.
- Hut, P. (2007). Virtual laboratories. *Progress of Theoretical Physics*, 164, 38-53.



- Hooper-Greenhill, E., Dodd, J., Moussouri, T., Jones, C., Pickford, C., Herman, C., ... Toon, R. (2003). *Measuring the outcomes and impact of learning in museums, archives and libraries*. Retrieved from the Research Centre for Museums and Galleries website: <http://hdl.handle.net/2381/65>
- Karoulis A., Sylaiou S., & White M. (2006). Usability evaluation of a virtual museum interface. *Informatica*, 17(3), 363-380.
- Ke, F., & Grabowski B. (2007). Gameplaying for math's learning: Cooperative or not? *British Journal of Educational Technology*, 38(2), 249-259.
- Kunkel, T., Averkiou, M., & Chrysanthou, Y. (2008). A web-based virtual museum application. In M. Ioannides, A. Addison, A. Georgopoulos, L. Kalisperis (Eds.), *Proceedings of the 14th International Conference on Virtual Systems and Multimedia* (pp. 275-277). Budapest, Hungary: ARCHAEOLOGIA
- Loscos, C., Tecchia, F., Frisoli, A., Carozzino, M., Widenfeld, H. R., Swapp, D., & Bergamasco, M. (2004). The museum of pure form: touching real statues in an immersive virtual museum. In D. W. Fellner & S. N. Spencer (Eds.), *Proceedings of the 5th International Symposium on Virtual Reality, Archaeology and Cultural Heritage* (pp. 271-279). The Eurographics Association.
- McLeod, J., & Kilpatrick, K. M. (2001). Exploring science at the museum. *Educational Leadership*, 58(7), 59-63.
- Nijholt, A. (2000). Agent-supported cooperative learning environments. In Kinshuk I., Jesshope C., & Okamoto T., *International Workshop on Advanced Learning Technologies (IWALT 2000)* (pp. 17-18). Los Alamitos, CA: IEEE Computer Society Press.
- Papastergiou, M. (2009). Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. *Computers & Education*, 52(1), 1-12.
- Patias, P., Chrysanthou, Y., Sylaiou, S., Georgiades, C., Michael D., & Stylianidis, S (2008). The development of an e-museum for contemporary arts. In M. Ioannides, A. Addison, A. Georgopoulos, & L. Kalisperis (Eds.), *Proceedings of the 14th International Conference on Virtual Systems and Multimedia*. Budapest: ARCHAEOLOGIA.
- Read, J. C. (2008). Validating the fun toolkit: An instrument for measuring children's opinions of technology. *Cognition Technology & Work*, 10(2), 119-128.
- Resnick, M. (2001). *Revolutionizing learning in the digital age*. Retrieved from <http://people.ku.edu/~ghada/portfolio/standards/st2/digital.pdf>
- Reynolds, R., Walker K., & Speight C. (2010). Web-based museum trails on PDAs for university-level design students: Design and evaluation. *Computers & Education*, 55(3), 994-1003.
- Roussou, M. (2001, June). *Immersive interactive virtual reality in the museum*. Paper presented at the Trends in Leisure Entertainment 2001, London, UK.
- Tüzün H., Yilmaz-Soylu M., Karakus T., Inal Y., & Kizilkaya, G. (2009). The effects of computer games on primary school students' achievement and motivation in geography learning. *Computers and Education*, 52(1), 68-77.
- Tzanavari, A., Vogiatzis, D., Zembylas, M., Retalis, S., & Lalos, P. (2005). Affective aspects of web museums. *Studies in Communication Science*, 5(1), 39-56.
- Vavoula G., Sharples M., Rudman P., Meek J., & Lonsdale P. (2009). Myartspace: Design and evaluation of support for learning with multimedia phones between classrooms and museums. *Computers & Education*, 53, 286-299.
- Wishart J., & Triggs P. (2010). MuseumScouts: Exploring how schools, museums and interactive technologies can work together to support learning. *Computers & Education*, 54(2), 669-678.
- Wojciechowski R., Walczak K., White M., & Cellary W. (2004). Building virtual and augmented reality museum exhibitions. In D. P. Brutzman, L. Chittaro, & R. Puk. (Eds.), *Proceedings of the ninth international conference on 3D Web technology* (pp. 135-144). New York, NY, USA : ACM.
- Wrzesien M., & Raya M. A. (2010). Learning in serious virtual worlds: Evaluation of learning effectiveness and appeal to students in the e-junior project. *Computers & Education*, 55(1), 178-187.

## Appendix

Questionnaire for user experience:

1) How was your experience in this visit?



Awful



Not very good



Good



Really good



Brilliant

2) Would you like to do it again?

Yes	Maybe	No

## How University Students Evaluate Online Information about a Socio-scientific Issue and the Relationship with their Epistemic Beliefs

Fang-Ying Yang<sup>1\*</sup>, Yu-Hsin Chen<sup>2</sup> and Meng-Jung Tsai<sup>3</sup>

<sup>1</sup>Graduate Institute of Science Education, National Taiwan Normal University, 88 Sec. 4, Ting-Zhou Road, Taipei 116, Taiwan // <sup>2</sup>Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, 43 Sec.4, Keelung Rd., Taipei, 106, Taiwan // <sup>3</sup>Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, 43 Sec.4, Keelung Rd., Taipei, 106, Taiwan // fangyang@ntnu.edu.tw // hsin0902@gmail.com // mjtsai99@mail.ntust.edu.tw

\*Corresponding author

(Submitted November 22, 2011; Revised May 12, 2012; Accepted July 10, 2012)

### ABSTRACT

The purpose of the study was to explore the judgment criteria used by university students for evaluating online information about a socio-scientific issue and the associations, if any, with their epistemic beliefs. The problem context was a socio-scientific issue concerning the impact of electromagnetic waves on human health. The participants were 36 university students, who were asked to read a news report, briefly state their thoughts, and then proceed to the web search activity in order to determine if their thoughts were legitimate. The students' search processes were recorded by web camera. After the search activity, all students were interviewed about the criteria used to determine the credibility of the online information. The students responded to the interview questions as they were watching their own web search processes. Epistemic beliefs were then assessed using questionnaires. A coding scheme was developed to analyze the students' oral responses. It was found that these university students determined the credibility of the online information mostly with reference to the richness and explanative power of argument, the presence of evidence, and the authority source of information. However, few of them went further to examine the validity of the evidence. Correlation analyses and ANOVA showed that the use of overall criteria for judging the online information was associated significantly with students' epistemic beliefs regarding authority. Meanwhile, the number of criteria used for justifying evidence was significantly associated with the students' beliefs about learning ability and justification in science.

### Keywords

Web-based learning, Online information credibility, Epistemic beliefs, Media and science education

### Introduction

In many work places, public areas, or at home, people rely on the World Wide Web to search for information that can help them solve problems and make decisions. Although online technology allows instant and easy access to knowledge and information, there are few constraints on what kinds of information should be posted on the Internet.

Consequently, web users need to develop effective judgment schemes to help them distinguish useful and credible online information. In the domain of science education, evaluating information is an essential part of a "scientific habit of mind," which emphasizes the critical and evaluative thinking about information as the evidence for any claim or theory. Nevertheless, until recent years, few studies in science education have paid attention to how students think and make judgments about the online information that concerns the application of science and technology. Thus, the purposes of this study were to explore the judgment schemes employed by university students for evaluating online information, and discuss the factors that might influence their evaluative behaviors.

### Literature review

#### Criteria for judging online information

In the literature, a considerable amount of research regarding information searching can be found in the marketing, medical and health, and communication fields. In these studies, web search activities are linked to problem-solving and decision-making processes where the research questions are largely concerned with search strategies and consumer behaviors or characteristics (e.g., Darley, Blankson, & Juethge, 2010; Lin & Chan, 2009; Case, 2002). In

the last ten years, the World Wide Web has become a popular open source of information that can be easily accessed. However, since web information is often posted without proper screening, the issue about how web users evaluate the credibility of online information receives an increasing amount of attention from information researchers.

To justify online information, information researchers propose five key criteria—accuracy, authority, objectivity, currency, and coverage—as a guideline for information credibility (Metzger, 2007). However, empirical studies show that information seekers in general do not spend all of their mental efforts on evaluating the content of the online information. They determine its credibility largely by considering the characteristics of source, the quality of the data, the reputation of the source, the site's presentation, or sometimes only the surface features, depending on the search context (e.g., Eysenbach & Kohler, 2002; Fogg, Soohoo, Danielson, Marable, Stanford, & Trauber, 2003; Walther & Burkell, 2002; Metzger, 2007).

In the context of learning, Metzger Flanagin, and Zwarun (2003), found that although college students used the Web extensively for searching for general and academic information, they seldom verified what they found. Kimsey and Cameron (2005) reported that college students lacked the critical thinking skills to evaluate the credibility of online information. Sundin and Francke (2009) showed that upper secondary school learners (around age 17 and 18) in a social science program displayed unsophisticated search skills, and also had a hard time evaluating the credibility of information they found on the Web. Julien and Barker (2009) also demonstrated a similar result with students in biology classes. Zuccala (2010) showed that when searching the online open access of research literature, users referred to science journalists as fact interpreters. Meanwhile, universities and scholars were considered more credible as information sources. Studies that have examined the practice of the five key criteria mentioned previously demonstrated that these criteria were not used frequently in students' responses, and on many occasions, only one or two were applied (Metzger, 2007; Julien & Barker, 2009). In sum, the studies reviewed above suggest that inadequate criterion systems are employed by learners for evaluating online information.

In the domain of science education, a major topic of discussion regarding Internet technology is how to incorporate the WWW source into instructional design. A popular instructional approach is the inclusion of online search activities as a part of the curriculum (Jonassen, Peck, & Wilson, 1999; Linn, Davis, & Bell, 2004; Relan & Gillani, 1997; Tsai, 2005). In general, it has been shown that such activities promote self-regulated learning and help learners to construct knowledge. More recently, to enhance search abilities and outcomes, a number of researchers have directed their attention to information needs, searching skills or strategies, and learner characteristics (Howard & Massanari, 2007; Jansen, Booth, & Smith, 2009; Liang & Tsai, 2010; Tsai, 2009; Hwang, Tsai, Tsai, & Tseng, 2008). Although the roles that the online search activities can play in promoting knowledge construction have been recognized by science educators, until recently few studies have examined how learners evaluate online information during a web search. A thorough examination of this issue can reveal students' reasoning modes, providing baseline information for curriculum design aimed at promoting evaluative and reflective thinking.

Although literature about information credibility has been accumulating in recent years, most available studies used self-report surveys or after-task interviews to assess students' evaluation performances. There are also studies confined to off-line environments (Mason, Boldrin, & Ariasi, 2010; Mason, Arisai, & Boldrin, 2011). The main problem of these methods is that they could not reflect the judgment criteria that students naturally employ during a particular search task. Thus, in this study, an attempt was made to address this issue. The first research question is: "During a naturalistic task, by what criteria did university students evaluate online information when they were exposed to a science-related issue that involves science and its application?"

### **Epistemic beliefs and its relevance to the evaluative intention**

In the previous section, we argued that web users and learners at different levels tend to not employ legitimate schemes for evaluating online information. Information researchers have proposed a checklist training approach for improving evaluative ability. Basically, the approach emphasizes the one-by-one use of the five key criteria mentioned previously to guide the evaluation process (Metzger, 2007). Although the approach is useful, we found that users' information needs (motivations), ability, and the problem contexts mediate the evaluation results (Fritch & Cromwell, 2001; Walther & Burkell, 2002; Dutta-Bergman, 2004; Metzger, 2007). Hence, we believe that an in-depth investigation of these factors will help clarify problems regarding how and why web users employ or do not employ criteria necessary for judging online information.

As mentioned, information research points out that user motivation and ability moderate the degree to which users evaluate online information. According to Metzger (2007), user motivation is related to personal or situational factors. Thus, other than the task demands, personal goals, values, and beliefs would determine whether an information searcher has found what he or she needs.

In this study, we propose an investigation of the belief construct. In particular, we are interested in the students' personal epistemic beliefs. Kitchener (1983) put forward that human cognition could be differentiated into three levels of activities at the cognitive, metacognitive, and epistemic levels. The epistemic level of cognitive activity involves personal reflections on or thinking about the epistemological assumptions about knowledge and knowing. Some researchers argue that personal beliefs about learning also constitute epistemological assumptions (Schommer, 1993). In the literature, different terminologies and models have been proposed to describe the above-mentioned personal reflections, such as personal epistemology, personal epistemological beliefs/perspectives, and epistemic beliefs. Though there are still disagreements among psychologists about the nature and definitions of personal epistemological assumptions, empirical studies have found that personal beliefs about knowledge, knowing, and learning seem to mediate thinking, decision-making and knowledge construction (Kitchener, 1983; King & Kitchener, 1994; Hofer & Pintrich, 1997; 2002; Schommer, 1993; Yang, 2005; Yang & Tsai, 2010).

In the context of web-based learning, many studies have found significant associations between epistemic beliefs and learning approaches as well as outcomes (Soloman, 2000; Hartley & Bendixen, 2001; Tsai, 2004; Yang & Tsai, 2008; Yang & Chang, 2009). Recent studies regarding web searching (Tsai, 2004; Tsai & Tsai, 2003) have indicated that learners' beliefs about the subject knowledge to be learned have a large effect on online search behaviors. Moreover, two recent studies found that college and high-school students were able to activate their epistemic beliefs when doing web searches about a controversial science issue (Mason, Boldrin, & Ariasi, 2010; Mason, Arisai, & Boldrin, 2011). Based on these studies, we believe that epistemic beliefs should also play a critical role in mediating the use of judgment criteria for evaluating online information. Thus, the second research question in the study is, "To what extent were the students' personal epistemic beliefs related to the judgment criteria used for evaluating the online information?"

## **Research method**

### **Subjects**

The participants of the study were 36 university students taking introductory courses in science education in two national universities in Taiwan. These students were aged from 20 to 25 and voluntarily took part in the study. Most of the subjects were science majors. Four were social science majors.

### **Assessments Interview with "meta-recall" technique for assessing the judgment criteria**

To study the students' search behavior, we used an information search platform called Meta-Analyzer, developed by Hwang and colleges (2008), to record online search performance. This online platform was able to record the total search time, navigating time, and websites that the web searchers visited and explored. To assess their judgment criteria, we developed an interview protocol coupled with a "meta-recall" technique for data collection. The meta-recall was a concept inspired by a previous work conducted by Tsai (2001), who assessed students' cognitive structure using the interview procedure coupled with a meta-listening technique. In most studies that employ the interview method, subjects are usually asked to think aloud about what they did during an activity after the activity has been completed. Such a procedure suffers from the fact that subjects might not be able to track their actions precisely because of the high cognitive demand on their memory. To reduce the memory load, we used an online camera, WebCam, to record all the user actions displayed on the computer screen during the search task. After the search task, participants took part in interviews during which they were shown their own search processes and were asked to comment on the credibility of the websites that they had visited or skipped as displayed in their own search records. With the aid of this meta-recall technique, the students were brought back to the moments when they were checking the online information. We believed that the students' responses would therefore reflect more closely the actual criteria they used for evaluating the online information.

## Tools for assessing epistemic beliefs

To explore different aspects of the students' epistemic cognition, two questionnaires of epistemological beliefs were employed in the study. One is the Revised Epistemology Questionnaire (REQ) modified from Schommer's Epistemology Questionnaire (EQ) (Schommer-Aikin, 2004). The other is the Scientific Epistemology Questionnaire (SEQ), originally developed by Conley et al. (2004). The REQ is a domain-general tool, while SEQ focuses specifically on the knowledge of science.

The 36-item REQ was adapted and reduced from the 63-item Schommer's Epistemological Questionnaire (EQ) for college students (Schommer, 1998). The REQ was developed by factor analysis with 350 university students in Taiwan. Four epistemological factors were abstracted from the factor analysis: (a) authority knowledge, with 6 items (Cronbach's  $\alpha = 0.52$ ), (b) certain knowledge, with 10 items (Cronbach's  $\alpha = 0.68$ ), (c) simple knowledge, with 11 items (Cronbach's  $\alpha = 0.65$ ), and (d) Innate Ability with 9 items (Cronbach's  $\alpha = 0.71$ ). According to Hofer and Pintrich (2002), the domain-general construct of personal epistemology suffers somewhat from ambiguous definitions and uncertainty about the core constituents. Therefore, the existing epistemology questionnaires developed based on the domain-general assumptions were usually not particularly reliable. However, Hatcher and Stepanski (1994) have claimed that, for social science studies, a Cronbach's alpha coefficient even as low as 0.55 can be recognized and accepted for statistical consideration. Accordingly, except for "authority knowledge," which was slightly below this requirement, the alpha values of these epistemic factors were statistically acceptable. The factor structure of REQ was similar to EBQ (32 items) developed by Quian and Alvermann (1995), who employed the previous version of EQ and the same analysis procedure.

The 26-item, five-point Likert-style SEQ questionnaire was developed by Conley and others (Conley, Pintrich, Vekiri, & Harrison, 2004). It also consists of four factors, namely source (five items, Cronbach's  $\alpha = 0.82$ ); certainty (six items, Cronbach's  $\alpha = 0.79$ ); development (six items, Cronbach's  $\alpha = 0.66$ ); and justification (9 items, Cronbach's  $\alpha = 0.76$ ). In brief, "source" concerns the belief in authority knowledge, "certainty" refers to the belief in the right answer, "development" measures the belief about science as an evolving and changing subject, and "justification" focuses on how individuals justify knowledge. The SEQ questionnaire was translated into Chinese and has been tested with Taiwanese samples (Liang & Tsai, 2010).

## Procedure

The study involved activities of decision-making and online information searching. The problem context was a socio-scientific issue, namely "whether the electromagnetic waves (EMW) emitted by a cellular phone base station would cause cancer." Participants were first asked to write down their prior beliefs about whether EMWs are harmful to human health, and then two news reports from different standpoints regarding the effects of EMWs emitted from cellular base stations were given to the participants to read. They were then asked to write down their preferred position. Afterward, the students were asked to search freely on the World Wide Web in order to decide if their thoughts were legitimate. There was no time limit for their search because it was our intention to create a naturalistic task environment. Each student's search process was screen-captured by the WebCam software. Once the information search activity stopped, the interviews coupled with the meta-recall technique, as described in the instrument section, were conducted right away to probe the students' thoughts about the issue, and their criteria for judging the credibility of the information on the Web. The questionnaire surveys for epistemic beliefs were administered after the interviews.

## Data analysis

### Probing students' judgment criteria

As mentioned, to probe students' criteria for judging the online information, their whole online search processes were recorded and then shown during the interviews. The students were asked one webpage after another as they were displayed in their records to explain why they visited a particular website or stayed on some particular online information, and whether and why they thought the particular source or information was reliable. The interviews

were tape-recorded and afterward, all the responses were transcribed and analyzed according to the content analysis procedure.

To analyze the students' responses, we developed a coding scheme. Initially, the five key criteria identified in the literature regarding the credibility of the online information were considered, including accuracy, authority, objectivity, currency, and coverage (or scope) (Metzger, 2007). However, when analyzing the students' responses, we found that they actually displayed more complicated judgment systems for distinguishing the online information. Two aspects of information frequently appeared in the student responses, namely the content and the source of the information.

As far as content was concerned, the students proposed different concerns about the online information, such as whether the description was clearly and logically stated, whether the information was up-to-date (consistent with the currency criterion), whether the information was given without taking sides (aligned to the objectivity criterion), whether there were numerical data stated in the information, and so forth. As a matter of fact, the content criterion in many ways also overlapped with the coverage criterion, defined by the information research.

When speaking about the source of the information, many students would check whether the information was from authority sources including professional organizations, books, science magazines (the authority criterion), or other sources open to the public. In addition to the content and source categories, quite a few students talked about whether the online information was consistent with their own experiences, beliefs, and prior understandings.

Consequently, a coding scheme for assessing the online judgment criterion was constructed. Table 1 shows the main evaluation categories (indicated by "aspect"), the associated requirement of each aspect, the major criterion, and examples of student responses. Five students' responses were randomly selected for the inter-coder agreement analysis after the final coding scheme was established. The agreement was higher than 93%. The slight difference was discussed and reconciled. A single coder then did the rest of the coding analysis.

*Table 1.* The categories, aspects, major criteria and student responses for evaluating the online information

Aspect	Requirement	Major criteria	Student responses (examples)
Content	Argument	Richness	<ul style="list-style-type: none"> <li>It provided detailed information about the standards for the emission of electromagnetic waves (such as boundary conditions).</li> </ul>
		Power of explanation	<ul style="list-style-type: none"> <li>The information contained explanative information.</li> <li>It explained the cause and effects.</li> </ul>
		Clear scope	<ul style="list-style-type: none"> <li>It provided well-defined definitions and conditions of the damage.</li> </ul>
		Clear reference	<ul style="list-style-type: none"> <li>It provided background information of who posted the information (such as occupation and academic background).</li> </ul>
		Objectivity	<ul style="list-style-type: none"> <li>It provided information of both sides.</li> </ul>
		Logical and structured statement	<ul style="list-style-type: none"> <li>The arguments were logically and clearly stated.</li> </ul>
	Evidence	Controlled study	<ul style="list-style-type: none"> <li>There was a controlled experiment/study.</li> </ul>
		Empirical data	<ul style="list-style-type: none"> <li>It provided numerical data in the text.</li> </ul>
		Repetition	<ul style="list-style-type: none"> <li>It was only a one-time human experiment. More experiments are needed.</li> </ul>
	Format	Science report	<ul style="list-style-type: none"> <li>It was a medical report.</li> </ul>
		Descriptions of real cases	<ul style="list-style-type: none"> <li>They showed real incidents.</li> </ul>
Source	Organization	Authority	<ul style="list-style-type: none"> <li>The information was provided by some research organization.</li> <li>It was from a global group or organization.</li> </ul>
	Expert	Authority	<ul style="list-style-type: none"> <li>The information was provided by a scientist.</li> <li>It was posted by people with a proper background.</li> </ul>
	Website	Public accessibility	<ul style="list-style-type: none"> <li>It was from a wiki.</li> </ul>

Document	Reliability	<ul style="list-style-type: none"> <li>English sites are more reliable.</li> </ul>
	Authority	<ul style="list-style-type: none"> <li>It was a book chapter.</li> <li>It was from a thesis.</li> </ul>
Personal domain	Belief/theory	<ul style="list-style-type: none"> <li>It was consistent with what I believe.</li> </ul>
	Prior understanding	<ul style="list-style-type: none"> <li>It was consistent with my understanding.</li> </ul>

### Analyzing personal epistemic perspectives and their associations with the use of judgment criteria

As mentioned in the instrument section, two questionnaires (REQ and SEQ) were used to assess the students' epistemic perspectives in both general and science domains. Both REQ and SEV have four factors. Thus, the mean scores of each factor were calculated and compared via descriptive and correlation analyses. Meanwhile, to cross-analyze the students' judgment criteria and their epistemic beliefs, correlation analyses and one-way ANOVA were performed.

## Result

### The overall web search

According to the meta-analyzer records, the average time of a web search was 14.76 minutes ( $SD = 7.75$ ) and the mean number of visited web sites was 10.09 ( $SD = 6.44$ ). Among the visited sites, non-science-related websites were visited more often (mean = 8.49,  $SD = 5.51$ ) than science-related sites (mean = 1.60,  $SD = 1.59$ ).

### The criteria used for judging the online information

The coding result of the students' judgment criteria is shown in Figures 1 to 4. Basically, as Table 1 reveals, three aspects of online information, namely content, source, and personal aspects, could be extracted from the participants' responses. In the content aspect, the web searchers would consider three requirements. The first is the argument requirement: "Were the arguments well presented?" The second is the evidence requirement: "Was there any empirical evidence?" The third was the format requirement: "What types of online information were more reliable?"

To satisfy the argument requirement, whether the information provided enough details and varieties of viewpoint (richness), and whether the information contained sufficient explanations of the cause and effects (explanative power) were the two major criteria frequently mentioned by the participants (higher than 30%). Some students (about 25%) paid attention to whether the information was well organized and logically stated, and whether the conditions for the facts to occur and the extent of the effects were explicitly described (scope). However, only about 22% of participants were concerned with whether the information was objectively expressed without taking sides (objectivity), and even fewer students (about 10%) reflected on whether the information was up-to-date (currency). The coding result is shown in Figure 1.

Since the problem context of the study was science-related, most students expressed the need for evidence. As presented in Figure 2 for the evidence request, most participants considered the presence of empirical data (mentioned by 66.7% of the participants). Nevertheless, less than half of the students (about 42%) went further to check if the data came from repeated studies or had been mentioned many times in other online sources. About 30% of the students mentioned that related claims should be drawn based on controlled studies. Some participants (22%) expected to see descriptions of how the studies were done. In short, for some university students, empirical data alone could not guarantee the validity of the online information. How the data were obtained and tested were also critical points to be considered. However, the percentage distributions showed that most students did not exercise multiple criteria to justify evidence.



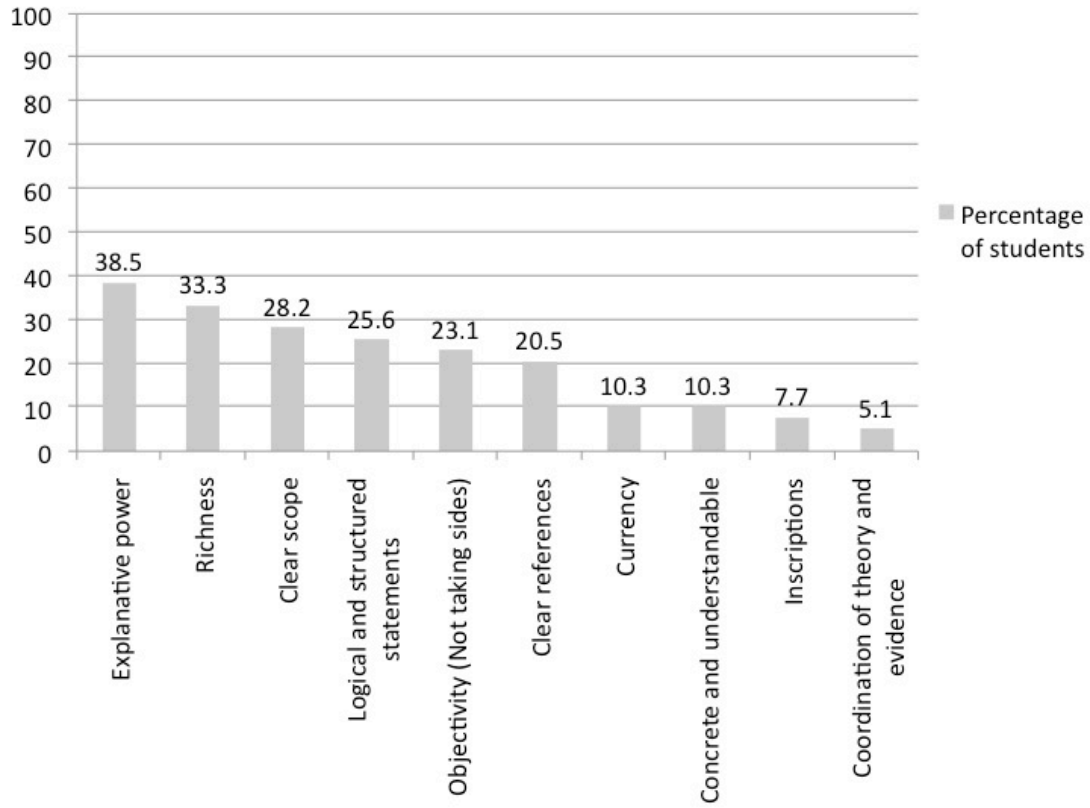


Figure 1. Percentage of students per argument criterion

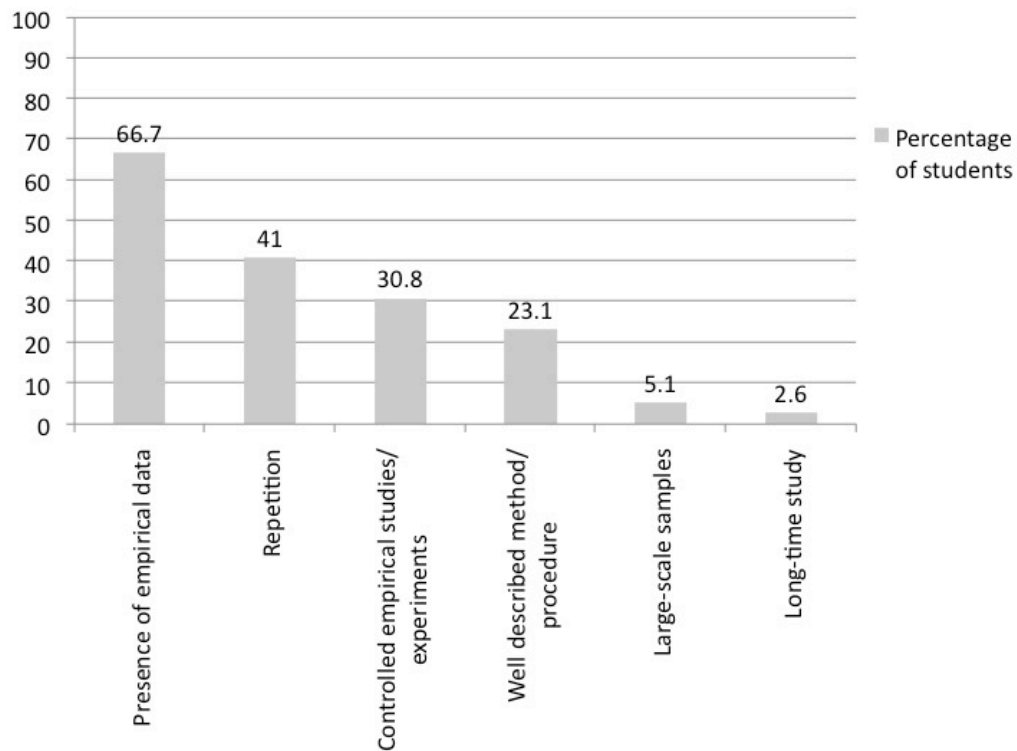


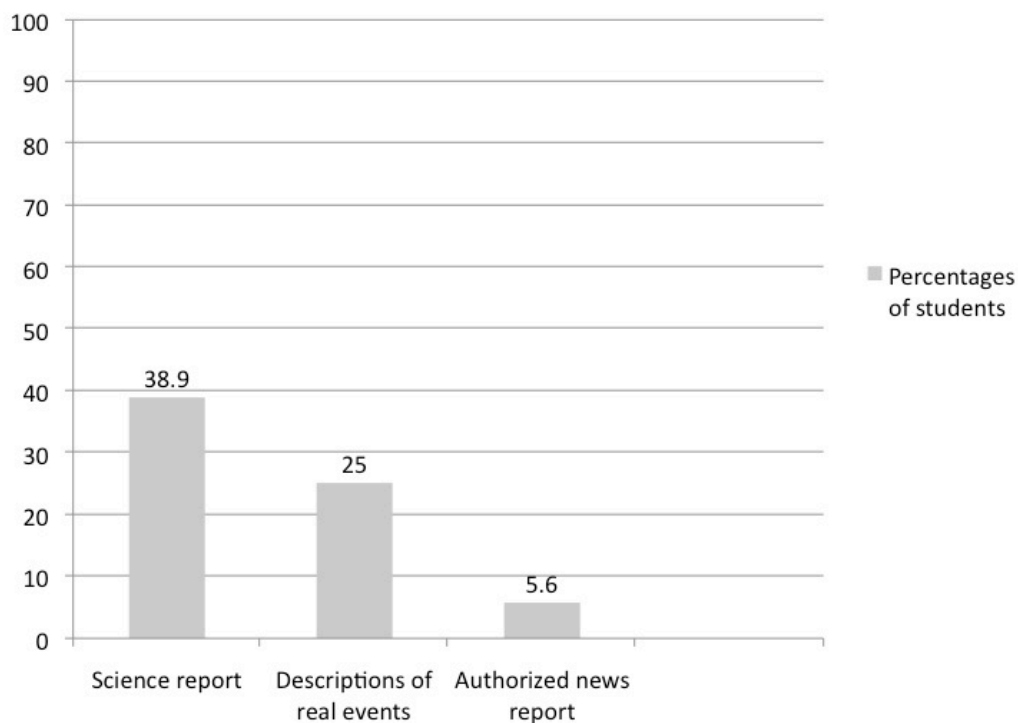
Figure 2. Percentage of students per evidence criterion

The frequency analysis of the criteria for “evidence” is shown in Table 3. We found that the majority of participants (about 86%) recognized the need for scientific evidence. Among these students, most were able to pinpoint one criterion, but only 36% of students went further to question the quality of the evidence.

*Table 3.* The numbers and percentages of “evidence” criteria

Number of criteria	Number of subjects	Percentage
0	5	13.9
1	18	50.0
2	8	22.2
3	4	11.1
5	1	2.8
Total	36	100.0

Furthermore, as indicated in the “format” requirement (see Figure 3), about 39% of the participants mentioned that the information would be more reliable if it was a science report or a description of real incidents.



*Figure 3.* Percentage of students per format criterion

From the students’ responses, we found that “source” was a vital aspect for them to look at. In general, as shown in Figure 4, information that came from authority (or professional) sources, such as professional organizations or expert figures, was considered to be trustworthy. In addition, if the information came from online sources that were open to the public, the degree of legitimacy might be increased. Finally, it was found that over 55% of students tended to take serious consideration of their own beliefs or understandings about the EMW issue when making judgments.

Students’ uses of these criteria in terms of mean numbers are displayed in Table 4. The mean number of overall criteria (referred to as CRITERIA in Table 4) used in judging the online information was 6.42. On average, the students recognized two aspects of the online information, and they were able to pinpoint four judgment requests. Meanwhile, they considered approximately four criteria on average in the content aspect and more than one in the source aspect. As for the personal aspect, over half of the subjects (55.6%) mentioned either their personal beliefs or prior understanding of the issue. The mean numbers of criteria of the two most frequently mentioned requirements in the content aspect are also listed in Table 4. They are “argument” and “evidence.” As noted, our subjects used

approximately two criteria to examine the arguments presented in the online information, but only 1.42 to evaluate evidence.

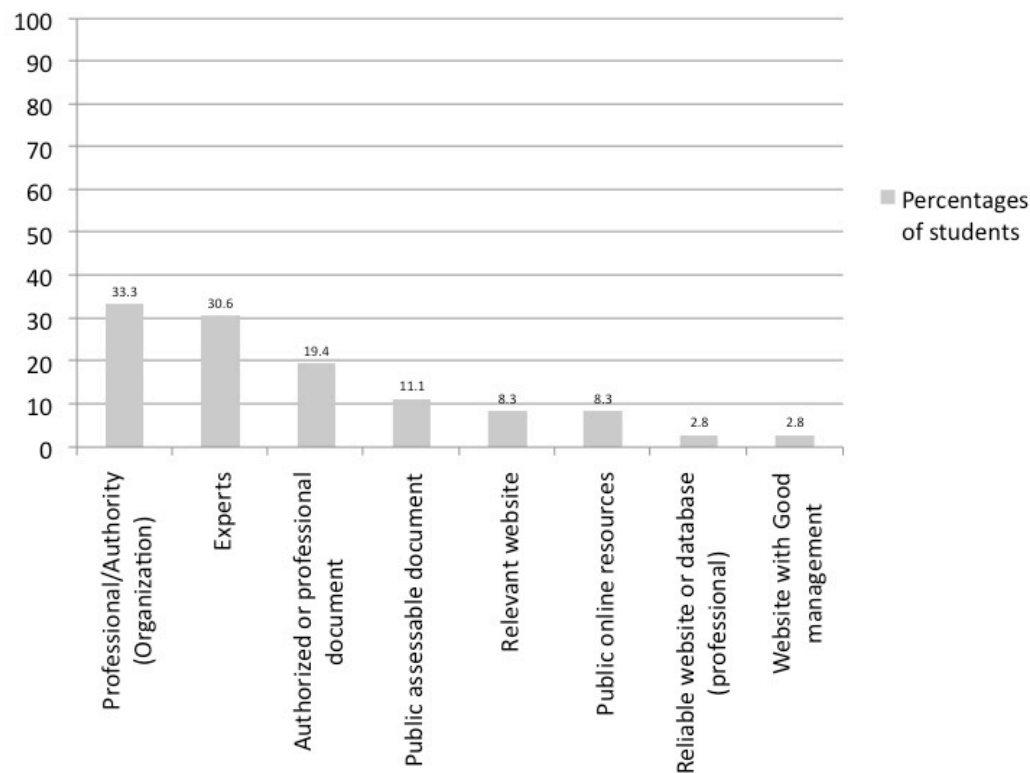


Figure 4. Percentage of students per format criterion

Table 4. The numbers of judgment criteria used for judging the online information

Criteria category	Minimum	Maximum	Mean	SD
ASPECT	1	3	2.31	0.62
REQUIREMENT	2	8	4.25	1.25
SOURCE <sup>1</sup>	0	5	1.50	1.28
CONTENT <sup>1</sup>	1	8	4.36	1.71
PERSONAL <sup>1</sup>	0	1	0.53	0.51
ARGUMENT <sup>2</sup>	0	5	2.11	1.26
EVIDENCE <sup>2</sup>	0	5	1.42	1.05
CRITERIA	3	12	6.42	2.38

Note: 1 = components of “aspects”; 2 = requirements of “content”

### Students’ epistemic beliefs

As mentioned above, two questionnaires, REQ and SEQ, were administered to assess the students’ epistemic beliefs. All participants ( $n = 36$ ) completed the REQ survey, but only 32 completed the SEQ. Tables 5 and 6 present the descriptive statistics for the REQ and SEQ results. According to Table 5, the university students in this study tended to agree that knowledge, in a general sense, comes from authority (mean = 3.04), but they thought more or less that knowledge is uncertain (mean = 2.77). On the other hand, when asked about scientific knowledge (as Table 6 shows), they seemed to have doubts about authority source (mean = 2.45) and believed to some extent that scientific knowledge is uncertain (mean = 2.65). Meanwhile, they agreed that scientific knowledge undergoes development and justification.

Table 5. The factor structure and scores of REQ (on a five-point Likert scale,  $n = 36$ )

Factor	Item	Reliability	Mean (SD)
AUTHORITY KNOWLEDGE (AK)	6	0.52	3.04 (0.51)
CERTAIN KNOWLEDGE (CK)	10	0.68	2.77 (0.53)
INNATE ABILITY (IA)	9	0.65	2.22 (0.42)
SIMPLE KNOWLEDGE (SK)	11	0.71	2.04 (0.53)

Table 6. The factor structure and scores of SEQ (on a five-point Likert scale,  $n = 32$ )

Factor	Item	Reliability	Mean (SD)
SOURCE (S)	5	0.81	2.45 (0.61)
CERTAINTY (C)	6	0.79	2.65 (0.72)
DEVELOPMENT (D)	6	0.58	4.42 (0.49)
JUSTIFICATION (J)	9	0.66	4.23 (0.42)

Table 7 reveals the associations between REQ and SEQ. As the table displays, beliefs in authority knowledge and certain knowledge in general (AK and CK) were significantly correlated to beliefs in authority sources and certain knowledge in science (S and C). Belief in the simple structure of knowledge in general (SK) was negatively correlated with beliefs in development and justification in science (D and J) to a rather high degree. In addition, belief in innate ability was moderately correlated with belief in certain knowledge in science and negatively correlated with belief in development and justification in science. In sum, epistemic beliefs regarding knowledge and learning in general were associated from moderate to high degrees with beliefs about the nature of scientific knowledge and the construction of scientific knowledge.

Table 7. Correlations between REQ and SEQ factors ( $n = 32$ )

	S	C	D	J
AK	0.43*	0.52**	-0.17	0.07
CK	0.34(*)	0.35*	-0.12	0.04
IA	0.21	0.38*	-0.35*	-0.31(*)
SK	-0.17	0.12	-0.47**	-0.63**

\* $p < 0.05$ ; \*\*  $p < 0.01$ ; (\*)  $p < 0.1$

### Associations between the use of judgment criteria and epistemic beliefs

To find out the associations, if any, between the use of judgment criteria and epistemic beliefs, we conducted a correlation analysis. According to the Spearman's rho analysis, the factor of authority knowledge in REQ was significantly correlated with the number of argument criteria ( $r = -0.39$ ,  $p < 0.05$ ) and approximately correlated with the total number of criteria used in judgment ( $r = -0.26$ ,  $p < 0.1$ ). As for SEQ, the factor of authority source was approximately associated with the number of argument criteria in the content aspect ( $r = -0.32$ ,  $p < 0.1$ ) and the number of overall criteria ( $r = -0.27$ ,  $p < 0.1$ ). In sum, the correlation analyses revealed that the number of criteria used in judging online information seemed to be more apparent in relation with beliefs in authority. It seems that the more students believed in authority, the fewer criteria they would apply to examine the content of online information.

In addition to the correlation analysis, an in-depth investigation was carried out by the one-way ANOVA. The criterion categories in Table 3 were included in the analysis. To perform one-way ANOVA, we grouped the students into three levels indicating high (H), mean (M), and low (L) criterion usage. The mean numbers of the criteria, as shown in Table 3, plus or minus 0.5 standard deviation, were used as the grouping standards. Since the number in each usage level was uneven, the test of homogeneity of variance was performed to examine the equality of the group variances. The significant findings reported here survived the test.

The ANOVA result showed that among these criterion categories, usage levels of the argument, and evidence criteria in the content aspects were statistically associated with the students' epistemic beliefs. Tables 8 and 9 demonstrate the findings.

Table 8. One-way ANOVA for epistemic beliefs and the use of evidence criteria

Epistemic beliefs	Usage level <sup>1</sup>	N	Mean	SD	ANOVA <sup>2</sup>	
					F	Post-hoc analysis
REQ Innate	L	5	2.00	0.29	3.57*	Level 2 > Level 3(*)
	M	18	2.36	0.43		
	H	13	2.03	0.33		
REQ Simple	L	5	1.64	0.35	3.00(*)	Level 2 > Level 1(*)
	M	18	2.23	0.59		
	H	13	1.98	0.41		
SEQ Development	L	4	4.92	0.10	3.02(*)	Level 1 > Level 2(*)
	M	15	4.28	0.42		
	H	13	4.40	0.56		
SEQ Justification	L	4	4.61	0.14	3.38*	Level 1 > Level 2 (*)
	M	15	4.07	0.42		
	H	13	4.32	0.40		

Notes:

1. Level L—no criterion used; level M—one criterion used; level H—more than one criterion used

2. \*  $p < 0.05$ ; (\*)  $p < 0.1$

Table 9. One-way ANOVA for epistemic beliefs and the use of the argument criteria

Epistemic beliefs	Usage level <sup>1</sup>	N	Mean	SD	ANOVA <sup>2</sup>	
					F	Post-hoc analysis
REQ Authority	L	11	3.26	0.29	4.34*	Level 1 > Level 3*
	M	12	3.14	0.43		Level 2 > Level 3 (*)
	H	13	2.74	0.33		
SEQ Authority source	L	9	2.76	0.88	2.40(*)	
	M	10	2.40	0.41		
	H	13	2.22	0.37		

Notes:

1. Level L—one criterion used; level M—two criteria used; level H—more than two criteria used

2. \*  $p < 0.05$ ; (\*)  $p < 0.1$

The post-hoc analyses displayed in Table 8 show that those in the high-usage level (that is, those who used more than one evidence criterion) to evaluate the online information tended to believe less in their innate ability. However, those who did not mention any evidence criteria (low usage) seemed to hold a stronger idea that knowledge is not just simple, piecemeal facts but has complicated structure and that scientific knowledge undergoes development and justification. As far as the use of argument criteria was concerned, the post-hoc analyses in Table 9 indicate that more criteria were used by students who believed less in authority.

## Discussion

### Use of judgment criteria

Our study demonstrates that when judging online information that involves scientific investigations, these university students activated complicated judgment systems to evaluate the credibility of the information. According to the content analysis, the students in this study paid rather high attention to the content and source aspects of the information. They examined the richness and explanative power of the arguments, required empirical evidence, and checked if the information came from authority sources. With respect to the five key criteria proposed by the information researchers, students in this study considered more the coverage and authority criteria but less the accuracy, objectivity, and currency criteria. This result was consistent with previous findings (e.g., Metzger, 2007; Julien & Barker, 2009). From the students' responses, it was also evident that they had employed various criteria beyond the scope of the five main criteria. These findings imply a lack of understanding of the information credibility of adult learners.

In the process of scientific argumentation, the inclusion of evidence is crucial for making a sound argument. However, while the evidence request was important for verifying the scientific claims, it is not explicitly included in the literature of information credibility. Even though, by definition, the evidence criterion might conceptually overlap the accuracy criterion, the former specifies the process of science. Hence, it is suggested that the evidence request, including the presence of empirical data and quality of evidence, should be regarded as a key criterion for the credibility of online information that involves scientific arguments. In this study, we found that most participants recognized the need for scientific evidence, but less than half went further to question the quality of the evidence. Such a finding echoes previous findings that even adults do not understand the role evidence plays in scientific argumentation (Kuhn, 1991).

### **The associations between students' epistemic beliefs and evaluation behaviors**

In this study, we also investigated a psychological factor that theoretically motivates the practice of judgment criteria, that is, the students' epistemic beliefs. One domain-general (REQ) and one domain-specific (SEQ) questionnaire were used to assess their beliefs. The correlation analysis found medium to high degrees of association between some of the factor scores of the two instruments, which indicate an overlap of the epistemic conceptions assessed by the two instruments.

The relationship between the students' epistemic beliefs and their use of judgment criteria was examined by correlation analysis and ANOVA. The correlation analyses showed that the more criteria used to examine the content of arguments, the less belief in authority the student had. The result of the one-way ANOVA also confirmed this finding. On the other hand, one-way ANOVA found that the use of criteria for verifying the presence and quality of evidence seemed to interact with beliefs in innate ability, the simplicity of knowledge, and development in science and justification in science. Further ad hoc analyses demonstrated that while students who exercised more evidence-based criteria tended to hold lower beliefs in their innate ability, some with lower scores in simple knowledge (indicating that they held the idea that knowledge is rather complex) mentioned no evidence at all. Similarly, students who gave higher scores for their beliefs about development and justification in science might not even recognize the role of evidence.

The above findings suggest complicated interactions between the use of evidence-related criteria and students' epistemic beliefs regarding learning ability, knowledge structure, and the construction of scientific knowledge. Although many students displayed advanced epistemic beliefs, not all of them recognized that evidence needs to be justified. In other words, even when students have seemingly developed advanced epistemic beliefs, they might still lack an in-depth understanding of the role of evidence in the development of scientific knowledge. The participants of this study could have regarded evidence as a part of scientific knowledge, and consequently found no need to justify it. This finding is consistent with previous studies showing that adults and adolescents tend to lack a proper understanding of the nature of evidence and theory (e.g., Kuhn, 1991; Yang, 2005). However, it should be noted that we conducted the correlation analysis and ANOVA in an attempt to probe possible relationships. Future studies should include large-scale samples or experimental designs to study in-depth the role of epistemic beliefs in the evaluation process.

### **Educational implications**

As discussed above, the participants of this study were able to practise complicated schemes to evaluate online information, but the schemes deviated from what information researchers have defined for information credibility. These findings suggest a need for training or learning activities aimed at promoting the understanding of information credibility. In particular, emphasis should be placed on accuracy, objectivity, and currency. As suggested by information researchers, a checklist approach based on the "big five" criteria (Metzger, 2007) would be a good way to focus web users' attention on the critical aspects of online information. Noticeably, considering the nature of the issue discussed in this study, which involved consideration of scientific information, the validity of evidence should also be included as a major criterion for information credibility. Accordingly, we recommend that the design of the training program for information credibility should take into consideration the problem context and the nature of information to be discussed.

In this study, we show that the use of criteria is associated with personal epistemic beliefs. This finding suggests that students' practices of judgment schemes could not be solely explained by unfamiliarity with information credibility. A thorough exploration of the nature of knowledge and the process of knowledge construction is critical for enhancing learners' intention to use the essential criteria mentioned in the study. It is thus recommended that, prior to the training program for information credibility, trainees should be given opportunities to reflect on their own personal views about the nature of knowledge and knowing. In this way, both trainer and trainees will have a chance to reach an agreement on the credibility criteria.

Another fact revealed by the study is that, after years of science education, university students majoring in science might still hold naïve views about scientific evidence. This finding is a warning for science educators because it implies that scientific knowledge could have still been portrayed in schools as an end product without the process of justification (Duschl, 1990). While there are many reform efforts to be made, teachers can make use of online resources to design curricula that emphasize scientific argumentation that allows learners to experience the process of knowledge construction in science (Duschl & Osborne, 2002). Course activities that include the searching and evaluation of online information related to controversial science issues or socio-scientific issues, such as the one presented in this study, could be an effective way to bring about discussions on the development of science and scientific knowledge.

### **Limitations and the future development of the study**

Although by way of the qualitative and quantitative analyses, this study made an attempt to reveal the criterion system employed by university students for evaluating online information, and how the use of criteria might be affected by students' epistemic beliefs, generalization of the study results should be made with caution. First, this study involved 36 university students who were mostly majoring in science. Considering the optimum requirements of quantitative analysis, the small number of subjects involved and the lack of students with non-science backgrounds would reduce the power of generalization. In terms of future research, a thorough and large-scale investigation with students from various academic backgrounds is recommended. Such a large-scale study will allow a further discussion of the prior knowledge effect. In addition, given that the current study analyzed students' evaluative criteria one by one, it is also expected that a large-scale study that applies cluster analysis may further reveal whether certain criteria are activated together. Second, since the study was mainly explorative in nature, whether a learning or training process can effectively change students' evaluative behavior as claimed in the previous section remains an open question. Thus, in the future, it is necessary to conduct experimental studies that aim to examine whether students' evaluation behaviors can be changed or guided by the instructional treatments.

### **Acknowledgements**

The funding of the work was supported by National Science Council in Taiwan under the grant numbers of NSC 95-2511-S-003-026-MY3, NSC 100-2631-S-003-006- and NSC 100-2511-S-003-039-MY3.

### **References**

- Case, D. O. (2002). *Looking for information: A survey of research on information seeking, needs, & behavior*. San Diego, CA: Academic Press.
- Conley, A. M., Pintrich, P. R., Vekiri, I., & Harrison, D. (2004). Changes in epistemological beliefs in elementary science students. *Contemporary Educational Psychology*, 29(2), 186–204.
- Duschl, R. (1990). *Restructuring science education*. NY, NY: Teachers College Press.
- Duschl, R., & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Studies in Science Education*, 38(1), 39–72.
- Darley, W. K., Blankson, C., & Juethge, D. J. (2010). Toward an integrated framework for online consumer behavior and decision making process: A review. *Psychology & Marketing*, 27(2), 94–116.

- Dutta-Bergman, M. J. (2004). The impact of completeness and web use motivation on the credibility of e-health information. *Journal of Communication*, 54(2), 253–269.
- Eysenbach, G., & Kohler, C. (2002). How do consumers search for and appraise health information on the World Wide Web? Qualitative study using focus groups, usability tests, and in-depth interviews. *British Medical Journal*, 324(7337), 573–577.
- Fogg, B. J., Soohoo, C., Danielson, D. R., Marable, L., Stanford, J., & Trauber, E. R. (2003, June). *How do users evaluate the credibility of web sites? A study with over 2,500 participants*. Paper presented at the Proceedings of the 2003 Conference on Designing for User Experiences, San Francisco, CA. Retrieved March 17, 2013, from <http://portal.acm.org/citation.cfm?doid=997078.997097>
- Fritch, J. W., & Cromwell, R. L. (2001). Evaluating internet resources: Identity, affiliation, and cognitive authority in a networked world. *Journal of the American Society for Information Science and Technology*, 52(6), 499–507.
- Hartley, K., & Bendixen, L. D. (2001). Educational research in the Internet age: Examining the role of individual characteristics. *Educational Researcher*, 30(9), 22–26.
- Hatcher, L., & Stepanski, E. J. (1994). *A step-by-step approach to using the SAS system for univariate and multivariate statistics*. Cary, NC: SAS Institute.
- Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, 67(1), 88–140.
- Hofer, B. K., & Pintrich, P. R. (Eds.). (2002). *Personal epistemology: The psychology of beliefs about knowledge and knowing*. Mahwah, New Jersey: Lawrence Erlbaum Associates
- Howard, P. N., & Massanari, A. (2007). Learning to search and searching to learn: Income, education, and experience online. *Journal of Computer-Mediated Communication*, 12(3), Art. 5.
- Hwang, G. J., Tsai, P. S., Tsai, C. C., & Tseng, J. C. R. (2008). A novel approach for assisting teachers in analyzing student web-searching behaviors. *Computers & Education*, 51(2), 926–938.
- Jansen, B. J., Booth, D., & Smith, B. (2009). Using the taxonomy of cognitive learning to model online search. *Information Processing & management*, 45(6), 643–663.
- Jonassen, D. H., Peck, K. L., & Wilson, B. G. (1999). *Learning with technology: A constructivist perspective*. New Jersey: Merrill.
- Julien, H., & Barker, S. (2009). How high-school students find and evaluate scientific information: A basis for information literacy skills development. *Library & Information Science Research*, 31(1), 12–17.
- Kimsey, M. B., & Cameron, S. I. (2005). Teaching and assessing information literacy in a geography program. *Journal of Geography*, 104(1), 17–23.
- King, P. M., & Kitchener, K. S. (1994). The development of reflective thinking in the college years: The mixed results. *New Directions for Higher Education*, 1993(84), 25–42.
- Kitchener, K. S. (1983). Cognition, metacognition and epistemic cognition. *Human Development*, 26(4), 222–232.
- Kuhn (1991). *The skill of argument*. Cambridge: Cambridge University Press.
- Liang, J. C., & Tsai, C.-C. (2010). Relational analysis of college science-major students' epistemological beliefs toward science and conceptions of learning science. *International Journal of Science Education*, 32(17), 2273–2289.
- Lin, J., & Chan, H. C. (2009). Understanding the beliefs and intentions in search and purchase functions in an e-commerce website. *IEEE Transactions on Engineering Management*, 56(1), 106–114.
- Linn, M. C., Davis, E. A., & Bell, P. (2004). *Internet environments for science education*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Mason, L., Boldrin, A., & Ariasi, N. (2010). Searching the web to learn about a controversial topic: Are students epistemically active? *Instructional Science*, 38(6), 607–633.
- Mason, L., Ariasi, N., & Boldrin, A. (2011). Epistemic beliefs in action: Spontaneous reflections about knowledge and knowing during online information searching and their influence on learning. *Learning and Instruction*, 21(1), 137–151.
- Metzger, M. J., Flanagin, A. J., & Zwarun, L. (2003). Student Internet use, perceptions of information credibility, and verification behavior. *Computers & Education*, 41(3), 271–290.



- Metzger, M. J. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology*, 58(13), 2078–2091.
- Quian, G., & Alvermann, D. (1995). Role of epistemological beliefs and learned helplessness in secondary school students. *Journal of Educational Psychology*, 87(2), 282–292.
- Relan A., & Gillani, B. B. (1997). Web-based instruction and traditional classroom: Similarities and differences. In B. H. Khan (ed.), *Web-based instruction* (pp. 41–46). Englewood Cliffs, NJ: Educational Technology Productions.
- Schommer, M. (1993). Epistemological development and academic performance among secondary students. *Journal of Educational Psychology*, 85(3), 406–411.
- Schommer, M. (1998). The influence of age and education on epistemological beliefs. *The British Journal of Educational Psychology*, 68(4), 551–562.
- Solomon D. L. (2000). Toward a post-modern agenda in instructional technology. *Educational Technology Research and Development*, 48(4), 5–20.
- Sundin, O., & Francke, H. (2009). In search of credibility: Pupils' information practices in learning environments. *Information Research: An International Electronic Journal*, 14(4), 418. Retrieved from <http://informationr.net/ir/14-4/paper418.html>
- Tsai, C.-C. (2001). Probing students' cognitive structures in science: The use of a flow map method coupled with a meta-listening technique. *Studies in Educational Evaluation*, 27, 257–268
- Tsai, C.-C. (2004). Information commitments in web-based learning environments. *Innovations in Education and Teaching International*, 41(4), 105–112.
- Tsai, M. J., & Tsai, C.-C. (2003). Information searching strategies in web-based science learning: The role of Internet self-efficacy. *Innovations in Education and Teaching International*, 40(1), 43–50.
- Tsai, C. C. (2005). Preferences toward Internet-based learning environments: High school students' perspectives for science learning. *Educational Technology & Society*, 8(2), 203–312.
- Tsai, M. J. (2009). Online information searching strategy inventory (OISSI): A quick version and a complete version. *Computers & Education*, 53(2), 473–483.
- Walthen, C. N., & Burkell, J. (2002). Believe it or not: Factors influencing credibility on the Web. *Journal of the American Society for Information Science and Technology*, 53(2), 134–144.
- Yang, F. Y., & Chang, C. C. (2009). Examining high-school students' preferences toward learning environments, personal beliefs and concept learning in web-based contexts. *Computers & Education*, 52(4), 848–857.
- Yang, F. Y. (2005). Student views concerning evidence and the expert in reasoning a socio-scientific issue and personal epistemology. *Educational Studies*, 31(1), 65–84.
- Yang, F. Y., & Tsai, C. C. (2010). Reasoning on the science-related uncertain issues and epistemological perspectives among children. *Instructional Science*, 38(4), 325–354.
- Zuccala, A. (2010). Open access and civic scientific information literacy. *Information Research: An Informational Electronic Journal*, 15(1). Retrieved from the Information Research website: <http://informationr.net/ir/15-1/paper426.html>

# The Effect of an Example-Based Dynamic Program Visualization Environment on Students' Programming Skills

Mehmet Tekdal

Department of Computer Education & Instructional Technologies, Faculty of Education, Çukurova University,  
01330 Sarıçam, Adana, Turkey // mtekdal@cu.edu.tr

(Submitted April 19, 2012; Revised November 02, 2012; Accepted January 14, 2013)

## ABSTRACT

The purpose of this study was to investigate and compare the effect of the two computer-assisted learning systems named the Example-Based Dynamic Program Visualization Environment (EDPVE) and the Example-Based Static Program Visualization Environment (ESPVE) on undergraduate students' programming skills in an introductory programming course. The study was conducted using a quasi-experimental, non-equivalent control group design with 81 pre-service teachers in a Turkish university. Two classes were randomly assigned to use EDPVE(experimental group) and ESPVE(control group) system. The results of the experiments showed that: (1) there were a significant difference between EDPVE and ESPVE groups in favour of the EDPVE group; and (2) there were no significant difference between male and female students in EDPVE and ESPVE groups respectively. The findings of this study suggest that teaching and learning programming can be improved through the use of dynamic program visualization techniques.

## Keywords

Interactive learning environments, Programming and programming languages, Gender studies, Program visualization, Computer-assisted learning

## Introduction

Learning of programming can be a difficult and complicated process. One of the main reasons is (Miyadera, Kurasawa, Nakamura, Yonezawa, & Yokoyama, 2007) that students have difficulty in understanding what a computer is actually doing when executing a program lines and what is simultaneously happening in memory. Program visualization and animation tools have been developed in order to present the execution of program lines using graphical effects. These tools are used to help students to better understand program codes and how the execution of a program works.

Program visualization (PV) is “a research area that studies ways of visually assisting learners in understanding behaviour of programs” (Rajala, Laakso, Kaila, & Salakoski, 2008). Diehl (2007) divided PV into dynamic program visualization (DPV) and static program visualization (SPV), in reference to the behavior and the structure of the program respectively. DPV tools show the dynamic execution of programs for debugging, profiling, and also for understanding program run-time behavior. On the other hand, SPV tools are useful for producing textual representations such as pretty-printing in order to increase users' understanding of program code (Storey, Wong, & Müller, 2000).

A considerable amount of literature has been published on program visualization and animation tools developed for use in programming education environments (LaFollette, Korsh, & Sangwan, 2000; Ben-Bassat Levy, Ben-Ari, & Uronen, 2003; Moreno & Joy, 2007).

Byrne, Catrambone and Stasko (1999) conducted two experiments in order to examine whether animations of algorithms would help students to learn them more effectively. The results showed that one-way animations may aid learning as they encourage learners to predict the behavior of algorithms. These researchers concluded that instead of animation, prediction was the key factor for learning in their study.

Another study (Hundhausen & Brown, 2007) implemented an algorithm development and visualization model named “What You See Is What You Code” for novice learners, which taught programming using an imperative paradigm. An analysis of its usability and field study shows that their model helps novices to quickly identify and correct programming errors and to develop semantically corrected code.

Naps et al. (2003) surveyed the role of visualization and engagement in computer science education and found their respondents agreed that using visualization can help students.

Although Empirica Control (EC) is a visual programming platform designed primarily for technology education, it can be used to teach programming. Students can use it to produce graphical representations of programs on flowcharts. The results of a teaching experiment with 34 eighth-grade students indicated that EC is a useful tool for learning the principal elements of programming with minimal teaching effort (Lavonen, Meisalo, Lattu, & Sutinen, 2003).

Many studies were conducted using the Jeliot series (Eliot, Jeliot I, Jeliot 2000 and Jeliot 3) of program visualization applications (<http://cs.joensuu.fi/jeliot/description.php>). Ben-Bassat Levy, Ben-Ari and Uronen (2003) carried out an experiment with Jeliot 2000 in a year-long course. The results of the experiment show that the animation program helps students who have difficulty in learning the abstract concepts of computer science. Another experiment was conducted by Moreno and Joy (2007) using the Jeliot 3 animation tool that Jeliot 3 animations are hard for novice students but helped them to debug their programs and that using the program was easy. The findings of another qualitative case study with Jeliot 3 showed that animation programs can help students to learn programming basics and may increase their motivation (Sivula, 2005). Still, in another research by Hongwarittorn and Krairit (2010), Jeliot 3 was used as a learning tool. The experimental results showed that the use of Jeliot 3 could help students improve their learning performance in Java programming and obtain higher scores than those who did not use Jeliot3, but also it was found that the use of tool did not affect students' long-term attitudes toward programming.

A study by Lee and Wu (1999) was conducted in order to investigate the effects of debugging practices on programming skills. They developed a system called DebugIt:Loop for debugging errors in loop programs. The results of two experiments conducted with 26 college and 46 senior high school students enrolled in an introductory Pascal course showed that their model was effective in improving novice programmers' programming skills.

An online tutorial called VINCE, was developed by Rowe and Thorburn (2000) to teach C programming by showing program graphically. VINCE provides students with a step-by-step visualization of the execution of the program, shows the memory map and provides comments at every step. An evaluation study showed that VINCE improves students' learning of programming and that students also find it to be well-designed.

Many PV tools have been developed and used to teach programming recently, but vary widely in features and capabilities. Gomez-Albarran (2005) conducted an in-depth analysis of nearly 20 tools, selected those which were outstanding and classified them into four categories as follows: (1) tools that include a simple and reduced development environment (THETIS, AnimPascal, BlueJ and DrJava); (2) example-based environments (ELM-PE, ELM-ART, WebEx and Javy); (3) tools based on visualization and animation (ANIMAL, LEONARDO, XTANGO, POLKA, JHAVE, The Teaching Machine of C++ and the Jeliot family); and (4) simulation environments (the Karel family, JkarelRobot and Alice).

Using carefully prepared visual examples can help students to understand even more complicated and abstract concepts (Lee & Owens, 2000; Clark & Mayer, 2003). A number of example-based tools have been developed in order to teach programming using examples. These tools include ELM-PE, ELM-ART, WebEx and Javy. ELM-PE is an example-based intelligent programming environment for learning the LISP programming language (Brusilovsky & Weber, 1996). ELM-ART, which was based on ELM-PE, is an adaptive web-based tool for learning LISP (Weber & Brusilovsky 2001). WebEx is a web-based tool to enable teachers to teach programming using an example-based approach in heterogeneous classes (Brusilovsky, 2001). Javy is a teaching tool designed to simulate the structure of a Java Virtual Machine (JVM) and allows users to learn JVM and the Java language compilation process (Gomez-Martin, Gomez-Martin, & Gonzalez-Calero, 2003).

Some of the other most widely used PV tools are summarized below:

- **Bradman** is a low-level Program Visualization tool. Smith and Webb (2000) conducted an experiment using Bradman to tests the efficacy of tool in assisting novice programmers learn programming concepts using C in an introductory programming course. The results of the experiment showed that students who studied with this tool achieved more understanding of some programming concepts than those without access it.

- **DrJava** (<http://drjava.sourceforge.net/>) is a lightweight programming development environment for writing Java programs. It is designed primarily for students, providing an intuitive interface and the ability to interactively evaluate Java code and also includes powerful features for more advanced users. Moreover it is a pedagogic programming environment that enables students to focus on designing programs, rather than learning how to use the environment (Allen, Cartwright, & Stoler, 2002).
- **JIVE** (<http://www.cse.buffalo.edu/jive/>) is an interactive execution environment and provides a rich visualization of the execution of Java programs. JIVE can be used both as a debugger for incorrect programs as well as for gaining a deeper insight into the behavior of correct programs. It also serves as a pedagogic tool for teaching object-oriented programming, and has been successfully used in undergraduate and graduate programming courses (Lessa, Jayaraman, & Cxyz, 2010).
- **VILLE** is a language-independent program visualization tool for teaching novice programming. Some key features of the tool are defining and adding new languages, visualization row by row, flexible control of the visualization both forwards and backwards, breakpoints, code line explanation, the parallel view shows the program code execution simultaneously in two different programming languages, call stack, and publish examples. Rajala, Laakso, Kaila, and Salakoski (2008) conducted an experiment to evaluate ViLLE's effectiveness in learning basic programming concepts. In this experiment, while the control group used only a textual programming tutorial, the treatment group used the same material and in addition could execute the examples with ViLLE. They argue that program visualization, more specifically the ViLLE tool, enhances students' learning regardless of previous programming experience. In other words the tool benefits novice learners more than learners with some previous experience.
- **EduVisor** is a program visualization tool developed based on the four high-level design principles from theories of learning and the theory of perception (Moons & Backer, 2009). Findings of the experiment by Moons, and Backer (2012) showed that the students and teachers see the environment and the way it is used as an invaluable part of their education and the environment can help with understanding programming concepts that most students consider very difficult.
- **OGRE** (Object-oriented GRaphical Environment) – a program visualization tool that has been developed to specifically address these memory related issues by providing three-dimensional, interactive visualizations that aid higher-level understanding of a program's execution in memory. The most important difference between OGRE and the above tools are its 3D environment, in which all the main features of a running program can be visualized. The experiments by developer of the tool showed significant improvements in students' understanding of C++'s memory-related topics (Milne & Rowe, 2004).

Although there are many studies which have evaluated SPV and DPV tools separately in the literature, this is the first study to compare the effect of DPV and SPV on learning programming. In this study, two computer-assisted learning systems (EDPVE and ESPVE) developed to teach introductory programming. The purpose of this study was to investigate the effectiveness of instruction with EDPVE system in terms of students' achievement, retention and gender differences in comparison to ESPVE system. So, the research questions that guide this study were formulated as follows:

- Is there a significant difference between the mean scores of the students in the EDPVE and those in ESPVE group?
- Is there a significant difference between the mean scores of the male and female students in the EDPVE and those in ESPVE group respectively?

## Method

### Research design

This study was conducted using a quasi-experimental, non-equivalent group design with pretest and posttest control groups. The research design is shown in Table 1.

Table 1: The research design

Groups	Pretest	Treatment	Posttest	Delayed-test
EDPVE	O <sub>1</sub>	X <sub>1</sub>	O <sub>3</sub>	O <sub>5</sub>
ESPVE	O <sub>2</sub>	X <sub>2</sub>	O <sub>4</sub>	O <sub>6</sub>

O<sub>1</sub>, O<sub>2</sub>: Pre-tests

X<sub>1</sub>: Teaching with the EDPVE system

X<sub>2</sub>: Teaching with the ESPVE system

O<sub>3</sub>, O<sub>4</sub>: Post-tests

O<sub>5</sub>, O<sub>6</sub>: Delayed-tests

## Participants

The participants were 81 undergraduate second-year students of the Faculty of Education, Computer and Instructional Technologies at the Department of Education of Cukurova University in Adana, Turkey. The study was conducted during “Programming Languages I” course.

## Sampling

The convenience sampling technique was adapted and used to select participants for the study. Two classes were selected and randomly assigned to the EDPVE and ESPVE groups. The EDPVE group included 41 (22 males, 19 females) students and the ESPVE group included 40 (23 males, 17 females) students.

## Instruments

In order to assess the students’ achievement in the EDPVE and ESPVE groups, a Programming Achievement Test (PAT) was constructed using three experts in the subject area. The scope of the PAT includes: variables, control statements, loops and arrays. The PAT consists of 25 multiple-choice items. Each item had five alternative answers.

Before using the PAT in this study, the test was administrated to a group of students (N = 77) in order to test the reliability of the test instrument. An analysis of the items showed that two of the items were too easy. These items were slightly revised and 25 items were used. The inter-item consistency was calculated using the Kuder-Richardson formula 20 reliability coefficient (KR-20) and a reliability coefficient of 0.83 was obtained.

## Materials

Two computer-assisted learning systems (EDPVE and ESPVE) have been developed and are used to teach programming. The ESPVE group used the ESPVE teaching system with static visualization examples, whereas the EDPVE group used the EDPVE teaching system with dynamic visualization examples. Both systems used the same number of examples. A screenshot of the EDPVE system is shown in Figure 1.

The EDPVE system interface has six windows. These windows include Animation Control, Program Code, Flowchart, Variables, Console and Status Bar. With the Animation Control window, the user can start/stop/restart the program execution, run the program lines step-by-step or continuously and adjust the program execution speed. When the program execution starts, visualizations of the program line in the Program Code and the Flowchart windows are highlighted at the same time. Although both the Program Code and the Flowchart windows are active concurrently, the students have the option to close one window and observe the program execution in another. During program execution, comments about the current line are shown in the Status Bar and the values of variables are listed in the Variables window. The Console window appears on the screen during the execution of the input and output commands to interact with the user. In addition, the Console window simulates the console of programming language and provides a way to experiment with the execution of the program, as in a real programming language environment.

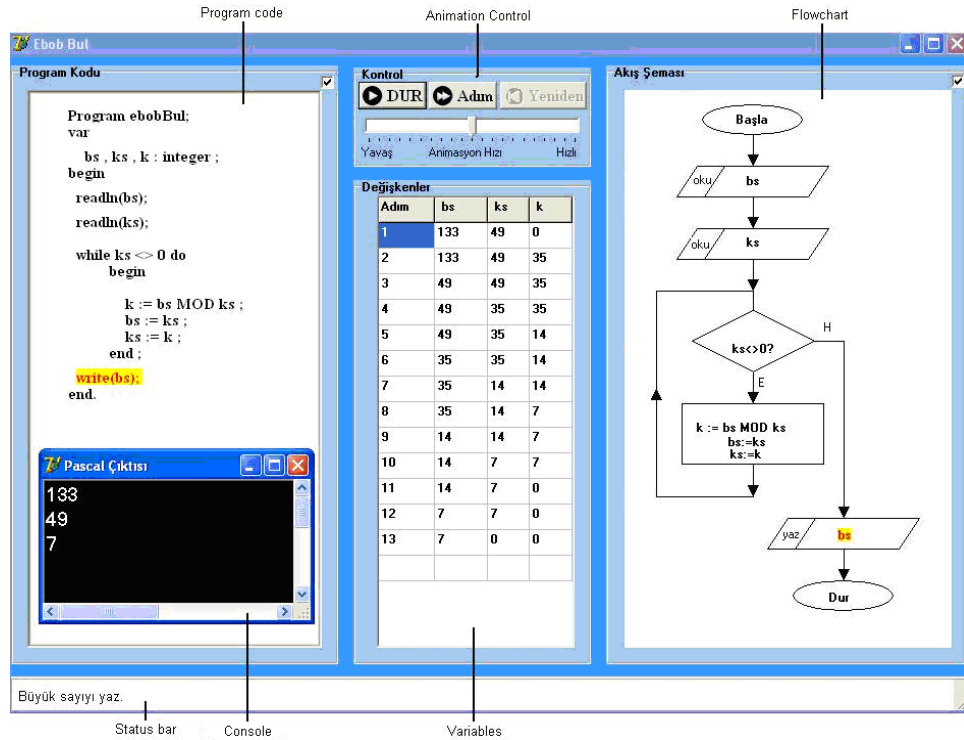


Figure 1. The Screenshot of the EDPVE system

The interface of the ESPVE system is similar to that of the EDPVE system, but it only has the Program Code, Variables and Flowchart windows. In this system, students can only inspect the program code and the related flowchart of static textual examples. A screenshot of the ESPVE system is shown in Figure 2.

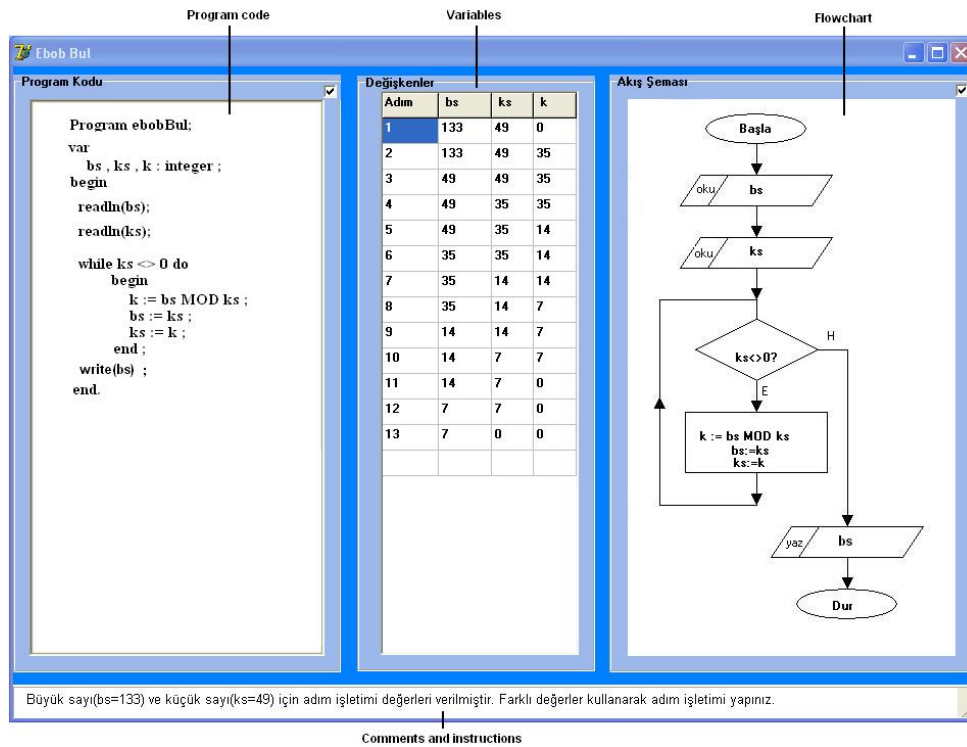


Figure 2. The Screenshot of the ESPVE system

## Procedure

The study was conducted during the students' normal lessons over 14 weeks. First, a pretest assessment was conducted with both groups before beginning the experiment. Second, the ESPVE group studied using traditional methods supported by the ESPVE system, while the EDPVE group was taught using traditional methods supported by the EDPVE system. Third, both groups completed their posttest assessment at the end of the experiment. Finally, three weeks after the posttest, both groups completed delayed-test assessments.

## Results

### Descriptive results

The sample sizes, means and standard deviations for the students' mean PAT scores were calculated. These values are listed in Table 2. A significance level of 0.05 was used throughout this study. Levene's test of homogeneity of variance was used in order to test the t-test assumption and showed that the variance between the groups was homogeneous. Non-parametric tests were conducted in order to test non-normally distributed data (Field, 2005).

Table 2. Sample sizes (N), means (Mean) and standard deviations (SD) of the students' PAT mean scores

Groups	N	Pretest		Posttest		Delayed-test	
		Mean	SD	Mean	SD	Mean	SD
EDPVE	41	10.51	2.52	21.41	2.93	21.05	2.99
ESPVE	40	10.78	2.53	19.15	3.63	18.13	3.14

### Learning results

Pretest, posttest and delayed-test significance levels are shown in Table 3. The results of the independent samples t-test analysis indicated that: (1) there was a non-significant difference between the pretest mean scores of the students in the EDPVE group ( $M = 10.51$ ,  $SD = 2.52$ ) and the ESPVE group ( $M = 10.78$ ,  $SD = 2.53$ ),  $t(79) = -.47$ ,  $p = .641$ ; (2) there was a significant difference between the posttest mean scores of the students in the EDPVE group ( $M = 21.41$ ,  $SD = 2.93$ ) and the ESPVE group ( $M = 19.15$ ,  $SD = 3.63$ ),  $t(79) = 3.09$ ,  $p = .003$ ; and (3) there was a significant difference between the delayed-test mean scores of the students in the EDPVE group ( $M = 21.05$ ,  $SD = 2.99$ ) and the ESPVE group ( $M = 18.13$ ,  $SD = 3.14$ ),  $t(79) = 4.29$ ,  $p = .000$ . In brief, pretest results indicate the EDPVE and ESPVE groups were equivalent and posttest and delayed-test results show significant differences favoring the EDPVE group.

Table 3. Pretest, posttest and delayed-test significance levels

Tests	Groups	N	Mean	SD	t	df	Significance(p)
Pretest	EDPVE	41	10.51	2.52	-.469	79	.641
	ESPVE	40	10.78	2.53			
Posttest	EDPVE	41	21.41	2.93	3.090	79	.003
	ESPVE	40	19.15	2.63			
Delayed-test	EDPVE	41	21.05	2.99	4.292	79	.000
	ESPVE	40	18.13	3.14			

### Retention results

Paired samples t-test results of pretest, posttest and delayed-test are shown in Table 4. The results of the paired samples t-test analysis indicated that: (1) there was a significant difference between the EDPVE group's pretest ( $M = 10.51$ ,  $SD = 2.52$ ) and posttest scores ( $M = 21.41$ ,  $SD = 2.93$ ),  $t(40) = -37.10$ ,  $p = .000$ ; (2) there was a significant difference between the ESPVE group's pretest ( $M = 10.78$ ,  $SD = 2.53$ ) and posttest scores ( $M = 19.15$ ,  $SD = 3.63$ ),  $t(39) = -30.53$ ,  $p = .000$ ; (3) there was a non-significant difference between the EDPVE group's posttest ( $M = 21.41$ ,  $SD = 2.93$ ) and delayed-test scores ( $M = 21.05$ ,  $SD = 2.99$ ),  $t(40) = .87$ ,  $p = .392$ ; and (4) there was a non-significant

difference between the ESPVE group's posttest ( $M = 19.15$ ,  $SD = 3.63$ ) and delayed-test scores ( $M = 18.13$ ,  $SD = 3.14$ ),  $t(39) = 1.29$ ,  $p = .204$ . In brief, the results indicated that there was not a significant difference between the posttest and delayed-test mean scores of both the EDPVE and ESPVE groups.

*Table 4. Paired samples t-test results of pretest, posttest and delayed-test significance levels*

Group	Test	N	Mean	SD	t	df	Significance(p)
EDPVE	Pretest	41	10.51	2.52	-37.10	40	0.000
	Posttest	41	21.41	2.93			
ESPVE	Pretest	40	10.78	2.53	-30.53	39	0.003
	Posttest	40	19.15	3.63			
EDPVE	Posttest	41	21.41	2.93	0.865	40	0.392
	Delayed-test	41	21.05	2.99			
ESPVE	Posttest	40	19.15	3.63	1.291	39	0.204
	Delayed-test	40	18.13	3.14			

### Gender results

The result of the independent samples t-test analysis indicated that: (1) there was a non-significant difference between the pretest mean scores of the male ( $M = 10.91$ ,  $SD = 2.31$ ) and female ( $M = 10.05$ ,  $SD = 2.74$ ) students in the EDPVE group,  $t(39) = 1.087$ ,  $p = .284$ ; (2) there was a non-significant difference between the posttest mean scores of the male ( $M = 22.05$ ,  $SD = 2.97$ ) and female ( $M = 20.68$ ,  $SD = 2.79$ ) students in the EDPVE group,  $t(39) = 1.506$ ,  $p = .140$ ; and (3) there was a non-significant difference between the delayed-test mean scores of the male ( $M = 21.27$ ,  $SD = 2.93$ ) and female ( $M = 20.79$ ,  $SD = 3.12$ ) students in the EDPVE group,  $t(39) = .511$ ,  $p = .612$ . The results are listed in Table 5.

The results of the independent samples t-test analysis indicated that: (1) there was a non-significant difference between the pretest mean scores of the male ( $M = 10.65$ ,  $SD = 2.65$ ) and female ( $M = 10.94$ ,  $SD = 2.41$ ) students in the ESPVE group,  $t(38) = -.354$ ,  $p = .726$ ; (2) there was a non-significant difference between the posttest mean scores of the male ( $M = 18.70$ ,  $SD = 3.61$ ) and female ( $M = 19.76$ ,  $SD = 3.68$ ) students in the ESPVE group,  $t(38) = -.918$ ,  $p = .364$ ; and (3) there was a non-significant difference between the delayed-test mean scores of the male ( $M = 18.22$ ,  $SD = 3.18$ ) and female ( $M = 18.00$ ,  $SD = 3.18$ ) students in the ESPVE group,  $t(38) = .214$ ,  $p = .832$ . The results are listed in Table 5.

In brief, the results indicated that there were no significant differences between males and females in both the EDPVE and ESPVE groups.

*Table 5. Results of pretest, posttest and delayed-test mean scores of males and females*

Groups	Tests	Gender	N	Mean	SD	t	df	Significance(p)
EDPVE	Pretest	Male	22	10.91	2.31	1.087	39	.284
		Female	19	10.05	2.74			
	Posttest	Male	22	22.05	2.97	1.506	39	.140
		Female	19	20.68	2.79			
	Delayed-test	Male	22	21.27	2.93	.511	39	.612
		Female	19	20.79	3.12			
ESPVE	Pretest	Male	23	10.65	2.65	-.354	38	.726
		Female	17	10.94	2.41			
	Posttest	Male	23	18.70	3.61	-.918	38	.364
		Female	17	19.76	3.68			
	Delayed-test	Male	23	18.22	3.17	.214	38	.832
		Female	17	18.00	3.18			

### Subtest results

For further analysis, in addition to a comparison of the two groups' PAT mean scores for the overall test items, the study also divided the PAT items into three subcategories: the Control Statements Test (CST)(7 items); the Loop



Statements Test (LST) (11 items); and the Array Statements Test (AST) (7 items), according to programming context subtopics. In this case, because the data could not meet the assumption of normality, the Mann-Whitney U-test was used instead of an independent samples t-test.

The results of the Mann-Whitney U-test analysis showed that: (1) there was a non-significant difference between the CST pretest mean scores of the students in the EDPVE group (Mdn = 4.00) and the ESPVE group (Mdn = 4.00),  $U = 757.50$ ,  $p = .515$ ; (2) there was a non-significant difference between the LST pretest mean scores of the students in the EDPVE group (Mdn=3.00) and the ESPVE group (Mdn = 3.00),  $U = 748.50$ ,  $p = .488$ ; and (3) there was a non-significant difference between the AST pretest mean scores of the students in the EDPVE group (Mdn = 3.00) and the ESPVE group (Mdn = 3.00),  $U = 814.50$ ,  $p = .957$ . The results are listed in Table 6.

The result of Mann-Whitney U test analysis showed that (1) there was a significant difference between the CST posttest mean scores of the students in EDPVE group (Mdn = 7.00) and ESPVE group (Mdn = 6.00),  $U = 588.00$ ,  $p = .020$ ; (2) there was a significant difference between the LST posttest mean scores of the students in EDPVE group (Mdn = 9.00) and ESPVE group (Mdn = 8.00),  $U = 551.00$ ,  $p = .010$ ; and (3) there was a significant difference between the AST posttest mean scores of the students in EDPVE group (Mdn = 7.00) and ESPVE group (Mdn = 6.00),  $U = 586.50$ ,  $p = .019$ . Results are listed in Table 6.

The results of the Mann-Whitney U-test analysis showed that: (1) there was a significant difference between the CST delayed-test mean scores of the students in the EDPVE group (Mdn = 7.00) and the ESPVE group (Mdn = 6.00),  $U = 530.00$ ,  $p = .003$ ; (2) there was a significant difference between the LST delayed-test mean scores of the students in the EDPVE group (Mdn = 8.00) and the ESPVE group (Mdn = 7.00),  $U = 549.00$ ,  $p = .010$ ; and (3) there was a significant difference between the AST delayed-test mean scores of the students in the EDPVE group (Mdn = 7.00) and the ESPVE group (Mdn = 5.50),  $U = 384.50$ ,  $p = .000$ . The results are listed in Table 6.

In brief, the results show pretest performance to be the same on each subcategory and significant differences in all subcategories on the posttest and delayed-test favoring the EDPVE group.

Table 6. Results of subcategories tests

Tests	Groups	N	Control Statements Test				Loop Statements Test				Array Statements Test			
			Mdn	U	Z	p	Mdn	U	Z	p	Mdn	U	Z	p
Pretest	EDPVE	41	4.0	757.5	-.650	.515	3.0	748.0	-.694	.488	3.0	814.5	-.053	.957
	ESPVE	40	4.0				3.0				3.0			
Posttest	EDPVE	41	7.0	588.0	-	.020	9.0	551.0	-	.010	7.0	586.5	-	.019
	ESPVE	40	6.0		2.328		8.0		2.572		6.0		2.353	
Delayed-test	EDPVE	41	7.0	530.0	-	.003	8.0	549.0	-	.010	7.0	384.5	-	.000
	ESPVE	40	6.0		2.929		7.0		2.583		5.5		4.297	

## Discussion and implications

The aim of this study was to investigate the effect of teaching/learning using the EDPVE system on students' programming skills. First, there was a significant difference between the posttest mean scores of the EDPVE and ESPVE groups in favour of the EDPVE group. This result in the achievement test suggests that the students in the EDPVE group had higher mean scores than the students in the ESPVE group, meaning that the learning effect of studying using the EDPVE system is superior to the effect of studying using the ESPVE system. The finding of the current study is consistent with a number of similar previous empirical researches (Hongwarittorn & Krairit, 2010; Lee & Wu, 1999; Smith & Webb, 2000; Rajala, Laakso, Kaila, & Salakoski, 2008; Milne & Rowe, 2004; Moons & Backer, 2009) on the use of PV techniques that improves the students learning performances. Although this study has provided a significant result, the findings of similar studies reported in the literature present both significant and non-significant results. The meta-study conducted by Hundhausen, Douglas, and Stasko (2002), which comprised an analysis of experimental studies of the effectiveness of algorithm visualization, showed that only 11 out of 24 experiments gave significant results regarding the effect of visualization on learning programming. Another meta-analysis conducted by Baker and Dwyer (2000) showed that visualization has a positive effect on students' achievement. One of the advantages of program visualization is that it provides a way to interact with programs and

to animate the execution of program code. Mayer and Moreno (2002) argue that animation has great potential to improve human learning by promoting a deeper understanding.

Second, there was a significant difference between the delayed-test mean scores of the EDPVE and ESPVE groups in favour of the EDPVE group. On the other hand, an analysis of the paired-sample t-test showed that there was not a significant difference between the posttest and delayed-test mean scores of both the EDPVE and ESPVE groups. This result indicates that studying under either the EDPVE system or the ESPVE system will have a retaining effect.

Third, the test results of CST, LST and AST showed that there was a significant difference between both the posttest and delayed-test mean scores of the EDPVE and ESPVE groups in favour of the EDPVE group. The result of the subtests are consistent with those of the overall test, and, once again, we can reasonably argue that learning effect of teaching/learning control, loop and array subtopics with the EDPVE system is better than the effect of using the ESPVE system. These findings of the current study are consistent with those of Ben-Bassat Levy, Ben-Ari, and Uronen (2003). In the case of 'if and when' statements, they found a significant improvement in the performance of the animation group, but no significant improvement with assignment or input/output statements. The present findings also seem to be similar to other research which has found that basic program structures, such as if, trigger and loop, were mastered (Lavonen, Meisalo, Lattu, & Sutinen, 2003).

Finally, no significant gender differences were found in this study. Although the posttest mean score of the male students (22.05) in the EDPVE group was better than that of the female students (20.68), and the posttest mean score of the female students (19.76) in the ESPVE group was better than that of the male students (18.70), neither the EDPVE nor the ESPVE group revealed significant differences. The findings of most of the previous studies have shown gender differences in programming achievement (Hess & Miura, 1983; Lieberman, 1985; Becker, 1986; Collis et al., 1988). However, the findings of the current study and a number of newer studies (Byrne & Lyons, 2001; Rountree, Rountree, & Robins, 2002; Bergin & Reilly, 2005; Lau & Yuen, 2009) have not found the expected gender differences. It is difficult to explain this result, but it may be related to the different learning environment which is provided for new generations. The students of today are different from those in earlier studies. Today's students have grown up in a rich technology-based teaching and learning environment, which incorporates information and communication technologies, meaning that they think, learn and visualize differently.

Although the findings of the present study indicate that EDPVE is effective for under-graduate students, further research is required in order to test whether it is effective for high school students. In addition, this study used a sample of students with no technical background. Therefore, further studies are being carried out in order to test whether similar results can be obtained with students who have scientific and engineering backgrounds.

The findings of this study have shown that the effect of dynamic program visualization has important implications for the future design and development of tools for teaching programming, and that it may be valuable to use dynamic program visualization techniques. This is an important issue for educators wishing to integrate dynamic program visualization techniques into their courseware.

### **Limitations and future works**

Due to limitations of the current system, students were used only a restricted number of pre-written examples of programming. For future studies, new versions of the learning system can be developed to help students not only use ready-to-use examples but also write and use her/his own examples. Based on the results of this study, other recommendations for future researches are (1) only quantitative research methods were used to compare the effect of learning systems on students' learning. In order to further explain why dynamic program visualization is more effective, it would be valuable to use both quantitative and qualitative research methods together depending on the opinions of students and teachers; (2) this study compared the two computer-assisted learning systems. It would be convenient to compare these systems with traditional teaching methods; and (3) this study also show that program visualization techniques can improve learning of programming. This potential can be adapted in new development teaching/learning environments.

## Acknowledgements

The author would like to thank to the editors and anonymous reviewers for their valuable comments on this manuscript.

## References

- Allen, E., Cartwright, R., & Stoler, B. (2002). DrJava: a lightweight pedagogic environment for Java. *Proceedings of the 33rd SIGCSE technical symposium on computer science education* (pp. 137–141). New York, NY: ACM. doi: 10.1145/563340.563395
- Baker, R. & Dwyer, F. (2000). A meta-analytic assessment of the effect of visualized instruction. *International Journal of Instructional Media*, 27(4), 417-426.
- Becker, H. (1986). Our national report card: Preliminary results from the new John Hopkins survey. *Classroom Computer Learning*, 6(4), 30-33.
- Bergin, S., & Reilly, R. (2005). Programming: Factors that influence success. *SIGCSE Bulletin*. 37(1), 411–415.
- Brusilovsky, P. (2001). WebEx: Learning from examples in a programming course. In W. Fowler, & J. Hasebrook (Eds), *Proceedings of WebNet'2001, World Conference of the WWW and Internet* (pp. 124-129). Retrieved from <http://www2.sis.pitt.edu/~peterb/papers/WebNet01.html>
- Brusilovsky, P., & Weber, G. (1996). Collaborative example selection in an intelligent example-based programming environment. In D. C. Edelson & E. A. Domeshek (Eds), *Proceedings of International Conference on Learning Sciences* (pp. 357-362). Retrieved May, 7, 2010 from <http://www2.sis.pitt.edu/~peterb/papers/icls96.html>
- Byrne, M. D., Catrambone, R., & Stasko, J. T. (1999). Evaluating animations as student aids in learning computer algorithms. *Computers & Education*. 33(4), 253-278.
- Byrne, P., & Lyons, G. (2001). The effect of student attributes on success in programming. *Proceedings of the 6th Annual Conference on Innovation and Technology in Computer Science Education* (pp. 49–52). New York, NY: ACM Press. doi: 10.1145/377435.377467
- Clark, R. C. & Mayer, R. E. (2003). *E-learning and the science of instruction*. San Francisco, CA: Jossey-Bass
- Collis, B., Therrien, D., Kass, H., Kieren, T., & Wood, P. (1988, June). *Gender differences in Canadian grade 11 students interactions with computers: Results of a national study*. Paper presented at the Annual Meeting of the Canadian Society for Studies in Education, Windsor, Ontario, Canada.
- Diehl, S. (2007). *Software visualization: Visualizing the structure, behaviour, and evolution of software*. Berlin, German: Springer.
- Field, A. (2005). *Discovering statistics using SPSS (and sex, drugs and rock'n'roll)* (2nd ed). London, UK: Sage.
- Gomez-Albarran, M. (2005). The teaching and learning of programming: A survey of supporting software tools. *The Computer Journal*. 48(2) 130-144.
- Gomez-Martin, P. P, Gomez-Martin, M. A., & Gonzalez-Calero, P. A. (2003). Javy: Virtual environment for case-based teaching of Java virtual machine. *Lecture Notes on Computer Science*, 2773, 906–913.
- Hess, R., & Miura, I. (1983). Gender and socioeconomic differences in enrollment in computer camps and classes. (Unpublished manuscript). Stanford University, Stanford, CA, USA.
- Hongwarittorn, N. & Krairit, D. (2010, April). *Effects of program visualization (Jeliot3) on students' performance and attitudes towards Java programming*. Paper presented at the spring 8th International conference on Computing, Communication and Control Technologies, Orlando, Florida, USA. Retrieved April 23, 2010, from [http://www.iiis.org/CDs2010/CD2010IMC/CCCT\\_2010/PapersPdf/TA750PM.pdf](http://www.iiis.org/CDs2010/CD2010IMC/CCCT_2010/PapersPdf/TA750PM.pdf)
- Hundhausen, C. D., Douglas, S. A., & Stasko, J. D. (2002). A meta-study of algorithm visualization effectiveness. *Journal of Visual Languages and Computing*. 13(3), 259-290.
- Hundhausen, C. D., & Brown, J. L. (2007). What you see is what you code: A “live” algorithm development and visualization environment for novice learners. *Journal of Visual Languages and Computing*, 18(1), 22-47.
- LaFollette, P., Korsh, J., & Sangwan, R. (2000). A visual interface for effortless animation of C/C++ programs. *Journal of Visual Languages and Programming*, 11(1), 27-48.

- Lavonen, J. M., Meisalo, V. P., Lattu, M., & Sutinen, E. (2003). Concretising the programming task: A case study in a secondary school. *Computers & Education*, 40(2), 115-135.
- Lau, W. W. F., & Yuen, A. H. K. (2009). Exploring the effects of gender and learning styles on computer programming performance: Implications for programming pedagogy. *British Journal of Educational Technology*, 40(4), 696-712.
- Lee, G. C., & Wu, J. C. (1999). Debug it: A debugging practicing system. *Computers & Education*, 32(2), 165-179.
- Lee, W. W., & Owens, D. L. (2000). *Multimedia-based instructional design: Computer-based training, web-based training, distance broadcast training*. San Francisco, CA: Jossey-Bass.
- Lessa, D., Jayaraman, B., & Czyz, J. K. (2010). *JIVE: A Pedagogic Tool for Visualizing the Execution of Java Programs*. (Technical Report 2010-13). Retrieved May 15, 2010, from <http://www.cse.buffalo.edu/tech-reports/2010-13.pdf>
- Ben-Bassat Levy, R., Ben-Ari, M., & Uronen, P. A. (2003). The Jeliot 2000 program animation system. *Computers & Education*, 40(1), 1-15.
- Lieberman, D. (1985). Research on children and microcomputers: A review of utilization and effects studies. In M. Chen & W. Paisley (Eds.), *Children and microcomputers: Research on the newest medium* (pp. 59-83). Beverly Hills, CA: Sage Publications.
- Mayer, R. E., & Moreno, R. (2002). Animation as an aid to multimedia learning. *Educational Psychology Review*, 14(1), 87-99.
- Milne, I., & Rowe, G. (2004). OGRE: Three-dimensional program visualization for novice programmers. *Education and Information Technologies*, 9(3), 219-237.
- Miyadera, Y., Kurasawa, K., Nakamura, S., Yonezawa, N., & Yokoyama, S. (2007). A real-time monitoring system for programming education using a generator of program animation systems. *Journal of Computers*, 2(3), 12-20.
- Moons, J., & Backer C. D. (2012). The design and pilot evaluation of an interactive learning environment for introductory programming influenced by cognitive load theory and constructivism. *Computers & Education*, 60(1), 368-384.
- Moons, J., & Backer C. D. (2009). Rationale behind the design of the EduVisor software visualization component. *Electronic Notes in Theoretical Computer Science*, 224, 57-65.
- Moreno, A., & Joy, M. S. (2007). Jeliot 3 in a demanding educational setting. *Electronic Notes in Theoretical Computer Science*, 178, 51-59.
- Naps, T.L., Rößling, G., Almstrum, V., Dann, W., Fleischer, R., Hundhausen, C., ... Velázquez-Iturbide, J.A. (2003). Exploring the role of visualization and engagement in computer science education. *ACM SIGCSE Bulletin*, 35(2), 131-152.
- Rajala, T., Laakso, M.-J., Kaila, E., & Salakoski, T. (2008). Effectiveness of program visualization: A case study with the ViLLE tool. *Journal of Information Technology Education: Innovations in Practice*, 7, 15-32.
- Rountree, N., Rountree, J., & Robins, A. (2002). Predictors of success and failure in a CS1 course. *SIGCSE Bulletin*, 34(4), 121-124.
- Rowe, G., & Thorburn, G. (2000). VINCE – an on-line tutorial tool for teaching introductory programming. *British Journal of Educational Technology*, 31(4), 359-369.
- Sivula, K. (2005). A qualitative case study on the use of Jeliot 3. (Unpublished master's thesis). University of Joensuu, Department of Computer Science, Finland. Retrieved May 13, 2010, from [ftp.cs.joensuu.fi/pub/Theses/2005\\_MSc\\_Sivula\\_Kimmo.pdf](ftp.cs.joensuu.fi/pub/Theses/2005_MSc_Sivula_Kimmo.pdf)
- Smith, P. A., & Webb, G. I. (2000). The Efficacy of a Low-Level program visualisation tool for teaching programming concepts to novice C programmers. *Journal of Educational Computing Research*, 22(2), 187-215.
- Storey, M.-A. D., Wong, K., & Müller, H. A. (2000). How do program understanding tools affect how programmers understand programs? *Science of Computer Programming*, 36, 183-207.
- Weber, G., & Brusilovsky P. (2001). ELM-ART: An adaptive versatile system for web-based instruction. *International Journal of Artificial Intelligence in Education*, 12, 351-384.

## Computer Based Assessment Acceptance: A Cross-cultural Study in Greece and Mexico

Vasileios Terzis<sup>1\*</sup>, Christos N. Moridis<sup>1</sup>, Anastasios A. Economides<sup>1</sup> and Genaro Rebolledo Mendez<sup>2</sup>

<sup>1</sup>Information Systems Department, University of Macedonia, Egnatia Street 156, Thessaloniki, 54006, Hellas, Greece // <sup>2</sup>Av. Xalapa esq. Av. Avila Camacho S/N, Xalapa, Veracruz, 91020 // bterzis@otenet.gr // papaphilips@gmail.com // economid@uom.gr // grebolledo@uv.mx

\*Corresponding author

(Submitted March 20, 2012; Revised August 23, 2012; Accepted October 10, 2012)

### ABSTRACT

The user's acceptance of Computer Based Assessment (CBA) Systems is examined with the help of the Computer Based Assessment Acceptance Model (CBAAM) in the two different cultures of Greece and Mexico. The study was conducted by delivering the same CBA system to students of identical courses in Greece and Mexico. The research data were analyzed using Partial Least Squares (PLS). The results indicate that the CBAAM is valid for both countries in overall. However, there are some cultural differences. Greek students' behavioral intention is triggered mainly by Perceived Playfulness and Perceived Ease of Use, while Mexican students' behavioral intention is caused by Perceived Playfulness and Perceived Usefulness. This study is a first step towards a cross-cultural analysis regarding CBA's acceptance and use.

### Keywords

Computer based assessment, Technology acceptance, Cross-culture comparison, Hofstede's cultural dimension, Perceived playfulness

### Introduction

Learning Management Systems (LMS) became a valuable tool for teachers and learners worldwide. The increased use of LMS drove to globalized educational software such as Blackboard and Moodle, that help teachers to provide high quality LMS to their learners. Globalization can also be found in computer based test and assessments such as Test Of English as a Foreign Language (TOEFL), Graduate Management Admission Test (GMAT) and Graduate Record Examination (GRE). Many students around the world take these exams and are prepared using simulated Computer Based Assessment systems.

Computer Based Assessment (CBA) is an integral service which comes along with LMS or alone. It provides many advantages to teachers and learners. Learners can find CBA very useful because they can practice in any lesson or specific task they want in order to ameliorate their weaknesses and to evaluate their performance (Joosten-ten Brinke et al., 2007; Kaklauskas et al., 2010). On the other hand, educators savor other advantages such as: test security, reduction of time and cost, automation of records and distance learning/marketing (Gvozdenko & Chambers, 2007; Smith & Caputi, 2007).

Despite the increased use of CBA, many learners are against using CBAs (Frankola, 2000). Thus, Researchers are trying to identify the factors that affect learners to use CBA. Researchers, based on models regarding Information Technology (IT) acceptance, e.g., Technology Acceptance Model (TAM) (Davis, 1989), developed many models to explain learner's acceptance and intention to use learning management systems (LMS) or CBA. Regarding CBA, Computer Based Assessment Acceptance Model (CBAAM) (Figure 1) is a model that includes many important variables to explain learner's acceptance on CBA (Terzis & Economides, 2011).

However, CBA's globalization questioned the invariance of CBAAM in other cultures around the world. Cultural differences that exist among different countries may affect CBAAM's effectiveness or the factors that affect learner's intentions to use CBA.

Thus, this paper aims to examine possible differences in computer based assessment acceptance between different cultures by applying CBAAM to Greek and Mexican students.

Section 2 presents previous studies that shed light on CBA acceptance or cultural effect on IT acceptance. Section 3 describes the methodology. Section 4 provides the data analysis and the results. Finally, sections 5 and 6 discuss the results and present the conclusions of this study respectively.

## **Literature review**

### **Computer based assessment acceptance**

Technology Acceptance Model (TAM) is the first and the most dominant model regarding IT acceptance (Davis, 1989). Davis developed TAM based on the Theory of Reasoned Action (TRA) (Fishbein & Ajzen, 1975). Another model that explains user's intentions is the Theory of Planned Behaviour (TPB) (Ajzen, 1991). Taylor and Todd (1995) presented a hybrid model which combined TAM and TPB. Unified Theory of Acceptance and Use of Technology (UTUAT) came to integrate previous models regarding IT acceptance (Venkatesh, Morris, Davis, & F. D. Davis, 2003).

LMS and CBA acceptance studies such as CBAAM have adopted variables from these previous models. From TAM, CBAAM and other studies have adopted Perceived Usefulness (PU) and Perceived Ease of Use (PEOU) (Landry, Griffith & Hartman, 2006; Lee, 2008; Ong, Lai & Wang 2004; Ong & Lai, 2006; Padilla-Melendez, Garrido-Moreno & Del Aguila-Obra, 2008; Teo, 2009; Van Raaij & Schepers, 2008; Yi & Hwang, 2003). From Unified Theory of Acceptance and Use of Technology (UTUAT), CBAAM and other LMS acceptance studies used Facilitating Conditions (Teo, 2009; Teo, Lee, & Chai, 2008) or Social Influence in their research models (Van Raaij & Schepers, 2008; Wang, Wu & Wang, 2009).

Furthermore, CBAAM included variables which were found to be more relevant with the context of learning and assessment acceptance. It adopted Perceived Playfulness (Moon and Kim, 2001; Wang et al., 2009). It proposed firstly Goal Expectancy which is based on Self-Management of Learning (Wang et al., 2009) and secondly Content (Shee & Wang, 2008; Wang, 2003).

CBAAM suggests that user's intentions to use a CBA are defined by Perceived Playfulness and Perceived Ease of Use. Perceived Usefulness is significantly attributed by Goal Expectancy, Content, Social Influence and Perceived Ease of Use. Usefulness, Content, Ease of Use and Goal Expectancy explain Perceived Playfulness Furthermore, Perceived Ease of Use explained by Computer Self Efficacy and Facilitating Conditions (Figure 1).

### **User acceptance across cultures**

The need to clearly understand the individual acceptance drives many researchers to a cross cultural analysis regarding acceptance. Several previous studies added a cross-cultural dimension by comparing the efficiency of an acceptance model such as TAM or by adding variables that are distinguishing the cultures.

Cultural differences showed that the culture has an impact on IT acceptance (Straub, Keil & Brenner, 1997; Straub et al., 2002). Specifically, Straub (1997) found through a three-country study that TAM did not remain invariant across different countries. Another cross-cultural analysis was applied in the context of pre-services teachers regarding TAM (Teo, Luan & Sing, 2008). Cross-cultural application of the TAM was performed also in the global consumer acceptance of international web sites among Brazilians, Germans, and Taiwanese (Singh, Fassott, Chao & Hoffmann, 2006). Other researchers applied TAM to countries that have many cultural differences with west such as Saudi Arabia (Al-Gahtani, Hubona & Wang, 2007) and People's Republic of China (Huang, Lu & Wong, 2003). In addition, UTUAT was also used to examine cultural differences in IT acceptance (Taksa & Flomenbaum, 2009; Im, Hong, & Kang, 2011). Cross-cultural analysis regarding IT acceptance was applied also in other contexts such as prepayment metering systems (Bandyopadhyay, K. & Bandyopadhyay, S., 2008), and e-learning (Keller, Hrastinski & Carlsson, 2007).

On the other hand, some studies extended TAM with cultural value-dimensions (eg. Zakour, 2004; Li et al., 2009) based on previous researches (Hofstede & Bond, 1988; Hofstede, 2001; Trompenaars, Hampden-Turner, 1988; Hall,

1989) regarding cultural differences. Finally, Gaspay, Dardan, & Legorreta (2007) delivered a very useful meta-analysis regarding the implementation of Hofstede's cultural dimensions in IT research.

This paper explores the cultural effect on CBA and highlight possible differences by applying CBAAM to Greek and Mexican students. Greece was selected as a European country, while Mexico as a Latin American country. In addition, the cultural values of these two countries indicated that the analysis will provide interesting results regarding the differences and similarities between different cultures and their effects on CBA's acceptance (Hofstede, 2001). Particularly, regarding differences between countries, Greece has the highest value regarding Uncertainty Avoidance, while Mexico is twentieth among sixty six (66) countries. Moreover, Mexico has one of the highest values (fifth) regarding Power Distance, while Greece is fortieth. Another smaller difference between the two countries is their values regarding Masculinity. Mexico is eighth in this cultural dimension while Greece is twenty-third. On the contrary, the fourth cultural dimension "Individualism" is almost the same for the two countries. The aforementioned differences and similarities between Greece and Mexico will provide useful insights regarding the effect of cultural dimensions on CBA's acceptance.

Thus, the present research is the first at examining the influence of national culture on the acceptance and use of CBA.

## **Methodology**

### **Data collection**

The survey study was conducted at two universities in Greece and Mexico. In order to eliminate any other effect except cultural, the questionnaire was distributed to first year students that were attending similar courses in the two universities. The course was an introductory informatics course. Students were educated regarding general concepts of Information Technology and basic use of internet and word processing.

The CBA includes questions from this course. The participation in the CBA was voluntary. 117 first-year Greek students, 45 males (38%) and 72 females (62%), signed up and appeared to the procedure. The average age of Greek students was 19.2 (SD = 1.03). In addition, 51 first-year Mexican students, 19 males (37%) and 32 females (63%), participated to the procedure. The average age of Mexican students was 18.9 (SD = 1.05). Furthermore, from the mean and standard deviation of the Computer Self Efficacy variable, we are able to understand that students from both countries felt confident regarding their computer skills (Table 1).

The CBA was a summative test in order to prepare students regarding their final examination for the introductory informatics course. Therefore, it was delivered two weeks after the end of the courses in order to give to the students the opportunity to be prepared for the procedure. The procedure was administered in January 2011 for both countries.

The CBA has the same characteristics for both groups. The only difference was the questions' language. Mexican students used the CBA in Spanish, while Greek students in Greek.

The CBA system is very simple. The questions appearance was randomized. Each question was appeared with four possible answers and a "next" button, which allowed the students to move to the next question after he/she answered the present question first. The CBA was build in a Windows XP machine using JavaScript with Perl CGI on Apache web server with MySQL (Moridis & Economides, 2009).

After the end of the procedure, the student has to answer the questionnaire survey (CBAAM). CBAAM consists of 30 items in order to measure its 9 latent variables. We used the seven point Likert-type scale with 1 = "strongly disagree" to 7 = "strongly agree", in order to measure the items.

## **Research model and hypotheses**

The proposed research model (Figure 1), CBAAM, was applied to Greek and Mexican students in order to examine the acceptance and use of CBA (Terzis & Economides, 2011). This section describes the CBAAM's variables and hypotheses.

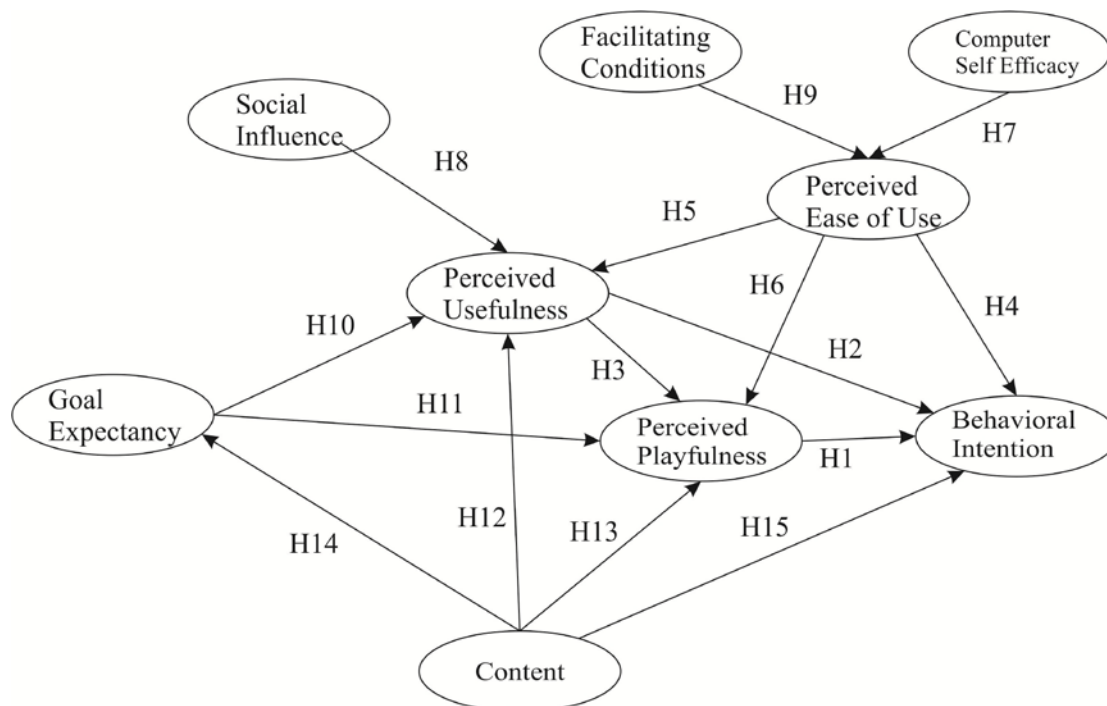


Figure 1. Research Model

### Perceived playfulness

Perceived Playfulness (PP) is an intrinsic significant belief that is formed from the individual's subjective experience with the system (Moon & Kim, 2001). Moon and Kim described PP as a variable defined by three dimensions such as concentration, curiosity and enjoyment. These dimensions are very essential for user's acceptance of a CBA. Previous studies showed that Perceived Playfulness is one of the most important determinants to use a CBA (Terzis & Economides, 2011).

Therefore we hypothesized:

H1: Greek and Mexican students' Perceived Playfulness will have a positive effect on their Behavioural Intention to use CBA.

### Perceived usefulness

The degree to which a user considers that using a specific system will improve his/her job performance is Perceived Usefulness (PU) (Davis, 1989). LMS acceptance models highlighted Perceived Usefulness as a very important factor of behavioral intention to use an e-learning system (e.g. Lee, 2008; Ong & Lai, 2006; Van Raaij & Schepers, 2008). Furthermore, Perceived Usefulness is also an essential determinant of Perceived Playfulness (Terzis & Economides, 2011). Therefore, we hypothesized:

H2: Greek and Mexican students' Perceived Usefulness will have a positive effect on their Behavioural Intention to use CBA.

H3: Greek and Mexican students' Perceived Usefulness will have a positive effect on their Perceived Playfulness.

### Perceived ease of use

The degree to which a user considers that using a system would be free of effort is Perceived Ease of Use (PEOU) (Davis, 1989). LMS acceptance models showed that the Perceived Ease of Use influences positively and directly the Perceived Playfulness, Perceived Usefulness and Behavioral Intention to Use (Venkatesh, 1999; Venkatesh & Davis, 1996).



H4: Greek and Mexican students' Perceived Ease of Use will have a positive effect on their Behavioural Intention to use CBA.

H5: Greek and Mexican students' Perceived Ease of Use will have a positive effect on their Perceived Usefulness.

H6: Greek and Mexican students' Perceived Ease of Use will have a positive effect on their Perceived Playfulness.

### **Computer self efficacy**

The degree to which a user perceives his/her capacity to use computers is Computer Self Efficacy (CSE) (Compeau & Higgins, 1995). CSE affects students during CBA. Individuals with higher CSE find the CBA system easier and consequently they have higher intentions to use the CBA system. The direct causal effect of CSE on PEOU and the indirect on Behavioral intention in the context of LMS or CBA acceptance are supported by previous studies (Agarwal, Sambamurthy, & Stair, 2000; Padilla-Melendez, Garrido-Moreno, & Del Aguila-Obra, 2008; Terzis & Economides, 2011).

H7: Greek and Mexican students' Computer Self Efficacy will have a positive effect on their Perceived Ease of Use.

### **Social influence**

Social Influence (SI) is a variable to measure the effect of other opinions on person behaviour and beliefs (Taylor & Todd, 1995). Social Influence consists of three dimensions: Subjective Norm, Image and Voluntariness (Karahanna & Straub, 1999). UTUAT combined the three dimensions and delivered Social Influence as one of the four key determinants of Behavioral Intention (Venkatesh et al., 2003). Social Influence was also found as a major determinant of behavioral intention in LMS and CBA contexts (e.g., Wang et al., 2009; Terzis & Economides, 2011).

H8: Social Influence will have a positive effect on Greek and Mexican students' Perceived Usefulness.

### **Facilitating conditions**

Facilitating Conditions (FC) are the system's characteristics or technical support that facilitate individuals to use a system. FC are delivered by help buttons, menus, Frequently Asked Questions (FAQ) and other facilities that may help users to interact with the system more effectively. Moreover, FC could be provided by the organization's staff. In CBA, tutors play an important role to proper delivery of the system to the students. So, we hypothesized a positive effect of FC on PEOU.

H9: Facilitating Conditions will have a positive effect on Greek and Mexican students' Perceived Ease of Use.

### **Goal expectancy**

Goal Expectancy (GE) is a construct that impacts an individual's belief that he/she is prepared properly to use the CBA system. GE has two dimensions. The first dimension is the student's beliefs regarding their preparation to answer the CBA's questions. The second dimension is the student's expectations and aspirations regarding the level of success. Previous studies showed that GE is a significant determinant of Perceived Usefulness and Perceived Playfulness (Terzis & Economides, 2011).

H10: Goal Expectancy will have a positive effect on Greek and Mexican students' Perceived Usefulness.

H11: Goal Expectancy will have a positive effect on Greek and Mexican students' Perceived Playfulness.

### **Content**

Content is a variable that measures user's perceptions regarding the CBA's corresponding course and CBA's questions. Course's content is an important factor for student's intentions to use the CBA. Students with higher

interest for a particular course will have higher intentions to use the corresponding CBA. In addition, the design of the CBA's questions is also essential regarding the user's beliefs and intentions. CBAs provided with questions that are higher in quality (understandable, relative with the course, innovative) and probably quantity are more likely to be used by the students. Therefore, we hypothesized:

H12: Content will have a positive effect on Greek and Mexican students' Perceived Usefulness.

H13: Content will have a positive effect on Greek and Mexican students' Perceived Playfulness.

H14: Content will have a positive effect on Greek and Mexican students' Goal Expectancy.

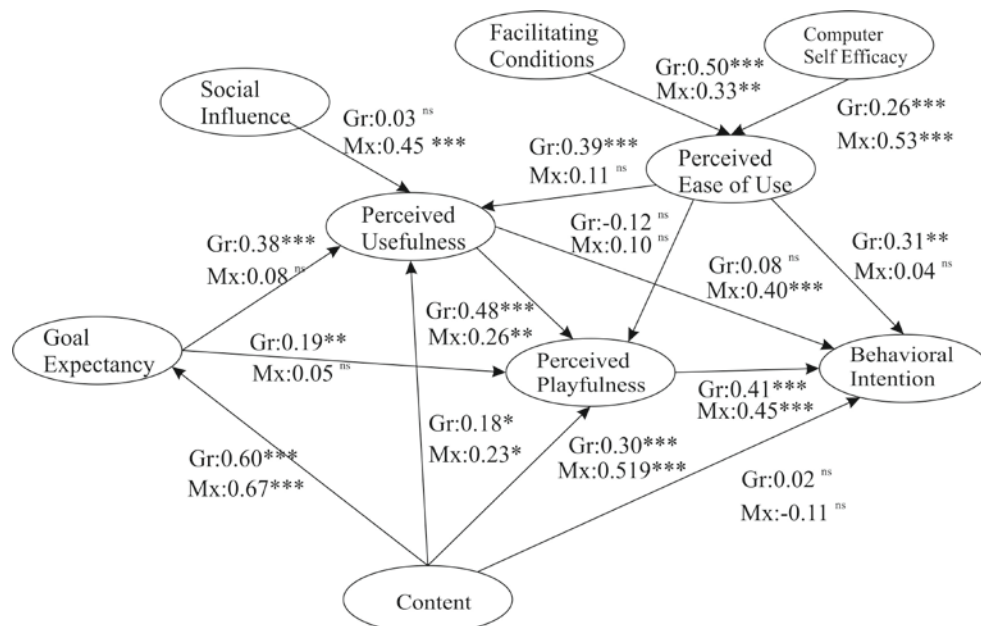
H15: Content will have a positive effect on Greek and Mexican students' Behavioral Intention to Use CBA.

Thus, this paper aims to examine the CBAAM's efficiency in different cultures by testing the aforementioned hypotheses in Greece and Mexico.

## Data analysis and results

Partial least-squares (PLS) analysis was used to measure the structural and the measurement model (Chin, 1998; Falk & Miller, 1992; Wold, 1992). The samples for both groups are large enough, since they surpass the minimum limit which is at least 10 times the largest number of independent variables impacting a dependent variable (Chin, 1998). In our model, the largest number of independent variables impacting a dependent variable is four (PU, PP, PEOU and C to BI). Thus, both the Greek and the Mexican samples are large enough since they are higher than 40.

Reliability and validity for the measurement model are measured through internal consistency, convergent validity and discriminant validity (Barclay, Higgins & Thompson, 1995; Wixon & Watson, 2001). Table 1 displays the items' reliabilities (Cronbach's alpha, Composite Reliability), Average Variance Extracted (AVE) and factor loadings, while Table 2 displays the variables' correlations and Square Roots of AVEs. The minimum values for these criteria are presented in Tables 1 and 2. Specifically: (1) Table 1 confirms convergent validity, (2) Tables 2 and 3 confirms discriminant validity for Greece and Mexico respectively. Thus, the results support the measurement model since they exceed the minimum values for both countries (Table 1, 2, 3). The only inconsistency is the second item regarding CSE which has factor loading lower than the minimum value (Table 1). Thus, CSE2 was omitted from the core construct regarding Mexico.



\* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ , ns: not significant.

Figure 2. Path coefficients of the research model.

The structural model is evaluated by examining the variance measured ( $R^2$ ) by the antecedent constructs. Values of the variance equal to 0.02, 0.13 and 0.26 are considered as small, medium and large respectively (Cohen, 1988). Secondly, t-values or p-values through the bootstrapping procedure are used in order to evaluate the significance of the path coefficients and total effects. Table 4 and figure 2 summarize the results for the hypotheses and  $R^2$ .  $R^2$  regarding Behavioral Intention and the most important model's variables such as Perceived Playfulness, Perceived Usefulness and Perceived Ease of Use are considered as large for both countries. In addition, most of the hypotheses are significant for both countries with some exemptions. The results are thoroughly discussed at the Discussions section.

Finally, Goodness of Fit (GoF) provides an overall prediction performance of the research model by taking into consideration the measurement and the structural models. The GoF is determined as the geometric mean of the average communality in the measurement model (AVE) and the average  $R^2$  of the endogenous variables (Tenenhaus, Vinzi, Chatelin, & Lauro, 2005). GoF is defined as small (0.10), medium (0.25) and large (0.36) (Wetzels, Odekerken-Schröder, & van Oppen, 2009). GoF is 0.60 and 0.61 for Greece and Mexico respectively. This means that the CBAAM is a reliable model to predict Behavioral Intention to use a CBA for both countries. The software for the data analysis was SmartPLS 2.0 (Ringle, Wende, & Will, 2005).

*Table 1. Results for the Measurement Model*

Construct Items	Mean		Standard Deviation		Factor Loading ( $> 0.7$ ) <sup>a</sup>		Cronbach $\alpha$ ( $> 0.7$ ) <sup>a</sup>		Composite Reliability ( $> 0.7$ ) <sup>a</sup>		Average variance extracted( $> 0.5$ ) <sup>a</sup>	
	GR	MX	GR	MX	GR	MX	GR	MX	GR	MX	GR	MX
Perceived Playfulness	4.95	5.34	1.15	0.84			0.86	0.80	0.91	0.87	0.71	0.62
PP1					0.77	0.80						
PP2					0.87	0.82						
PP3					0.86	0.77						
PP4					0.87	0.76						
Perceived Usefulness	5.09	5.24	1.18	1.04			0.88	0.88	0.93	0.92	0.82	0.80
PU1					0.90	0.91						
PU2					0.93	0.93						
PU3					0.88	0.85						
Perceived Ease of Use	5.56	5.55	1.15	0.98			0.77	0.72	0.87	0.84	0.69	0.64
PEOU1					0.84	0.88						
PEOU2					0.87	0.72						
PEOU3					0.77	0.78						
Computer Self Efficacy	5.33	5.91	1.00	0.82			0.84	0.7	0.89	0.82	0.67	0.61
CSE1					0.83	0.79						
CSE2					0.82	0.20						
CSE3					0.77	0.73						
CSE4					0.86	0.82						
Social Influence	5.81	4.57	0.96	1.01			0.83	0.77	0.89	0.85	0.66	0.59
SI1					0.78	0.70						
SI2					0.85	0.85						
SI3					0.81	0.80						
SI4					0.81	0.73						
Facilitating Conditions	5.96	5.20	0.91	1.02			0.80	0.70	0.91	0.87	0.83	0.76
FC1					0.89	0.91						
FC2					0.93	0.84						

Goal Expectancy	5.20	5.18	1.06	1.07		0.83	0.80	0.90	0.88	0.75	0.71
GE1					0.88	0.82					
GE2					0.86	0.87					
GE3					0.86	0.85					
Content	5.41	5.68	1.09	0.90		0.83	0.82	0.89	0.88	0.67	0.65
C1					0.87	0.86					
C2					0.85	0.77					
C3					0.73	0.78					
C4					0.80	0.82					
Behavioral Intention to Use	5.47	5.30	1.20	1.06		0.89	0.78	0.93	0.87	0.81	0.69
BI1					0.92	0.86					
BI2					0.88	0.74					
BI3					0.92	0.87					

a: Indicates an acceptable level of reliability and validity, GR: Greece, MX: Mexico.

Table 2. Discriminant validity for the measurement model in Greece

Constructs	BI	C	CSE	FC	GE	PEOU	PP	PU	SI
BI	<b>0.90</b>								
C	0.54	<b>0.82</b>							
CSE	0.24	0.32	<b>0.82</b>						
FC	0.34	0.49	0.17	<b>0.91</b>					
GE	0.41	0.60	0.41	0.23	<b>0.86</b>				
PEOU	0.57	0.61	0.34	0.54	0.40	<b>0.82</b>			
PP	0.63	0.65	0.34	0.30	0.63	0.46	<b>0.84</b>		
PU	0.60	0.65	0.31	0.33	0.65	0.66	0.72	<b>0.90</b>	
SI	0.48	0.50	0.27	0.54	0.44	0.48	0.41	0.46	<b>0.81</b>

Bold values: the square root of the average variance extracted (AVE) of each construct.

Table 3. Discriminant validity for the measurement model in Mexico

Constructs	BI	C	CSE	FC	GE	PEOU	PP	PU	SI
BI	<b>0.83</b>								
C	0.51	<b>0.81</b>							
CSE	0.16	0.42	<b>0.78</b>						
FC	0.18	0.56	0.36	<b>0.87</b>					
GE	0.44	0.67	0.53	0.45	<b>0.84</b>				
PEOU	0.46	0.69	0.65	0.52	0.59	<b>0.80</b>			
PP	0.65	0.78	0.33	0.54	0.60	0.63	<b>0.79</b>		
PU	0.65	0.61	0.39	0.32	0.51	0.55	0.66	<b>0.90</b>	
SI	0.44	0.55	0.38	0.52	0.47	0.51	0.63	0.67	<b>0.77</b>

Bold values: the square root of the average variance extracted (AVE) of each construct.

Table 4. Greece and Mexico results in relationships

		Greece (n= 117)			Mexico (n = 51)		
		$R_{Gr}^2$	$\beta_{Gr}$		$R_{Mx}^2$	$\beta_{Mx}$	
	BI	0.52			0.51		
	PP	0.60			0.68		
	PU	0.62			0.55		
	PEOU	0.36			0.51		
	GE	0.36			0.45		
H1	PP → BI		0.41***	support		0.45***	support
H2	PU → BI		0.08ns	not support		0.40***	support

H3	PU → PP		0.48***	support		0.26**	support
H4	PEOU → BI		0.31**	support		0.04ns	not support
H5	PEOU → PU		0.39***	support		0.11ns	not support
H6	PEOU → PP		-0.12ns	not support		0.10ns	not support
H7	CSE → PEOU		0.26***	support		0.53***	support
H8	SI → PU		0.03ns	not support		0.45***	support
H9	FC → PEOU		0.50***	support		0.33**	support
H10	GE → PU		0.38***	support		0.08ns	not support
H11	GE → PP		0.19**	support		0.05ns	not support
H12	C → PU		0.18*	support		0.23*	support
H13	C → PP		0.30***	support		0.52***	support
H14	C → GE		0.60***	support		0.67***	support
H15	C → BI		0.02ns	not support		-0.11ns	not support

\* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ , ns: not significant.

## Discussion

The good fit of the overall model in the two diverse cultures of Greece and Mexico indicates that the CBAAM could be used to explain user's behavior regarding CBA acceptance. The model was similar for both countries on the following effects: (1) Perceived Playfulness has the most important direct positive effect on Behavioral Intention to use the CBA. Previous studies also suggested Perceived Playfulness as a major determinant of Behavioral Intention to use an Information System or a Computer Based Assessment (Moon & Kim, 2001; Wang et al., 2009; Terzis & Economides, 2011), (2) Content has significant positive effect on Perceived Usefulness, on Perceived Playfulness and on Goal Expectancy but not direct significant effect on Behavioral Intention. Therefore, when a CBA's Content is designed wisely, students might experience CBA as useful and playful, and consequently it would be more likely to be used. Moreover, students' Goal Expectancy is affected by the CBA's Content. Course content and CBA's content have to be clear and comprehensible in order to facilitate students regarding their study, and goal estimation. (3) Perceived Usefulness has a significant direct effect on Perceived Playfulness. This means that the level of Perceived Usefulness determines the level of Perceived Playfulness. As we describe Playfulness is a three-dimensional construct which includes concentration, curiosity and enjoyment. Results showed that students' perceptions regarding usefulness for both countries affect positively their perceptions regarding the level of their concentration, curiosity and enjoyment during the CBA. Consequently, it might be assumed that the students use Playfulness to connect Usefulness with Behavioural Intention to Use. (4) Perceived Ease of Use is significantly attributed to Computer Self Efficacy and to Facilitating Conditions. The effect of Computer Self Efficacy indicates that a student who knows how to use computers, probably he/she will find easy to use a Computer Based Assessment that requires basic information technology skills. On the other hand, the effect of Facilitating Conditions suggests that if designers provide technical support through the system and tutors physical support with their presence to answer the students' queries, it is more likely students to find CBA ease of use. (5) On the other hand, the effect of Perceived Ease of Use on Perceived Playfulness was not significant for both countries. This means that the ease of use of the system does not provide any advantage in order to enhance their perceptions regarding the playfulness of the system. This result might be explained by the fact that students from both countries have very large Computer Self Efficacy (Table 1). Therefore, since students are qualified to use basic computer software, ease of use of the system could not affect students' perceptions regarding the playfulness of the system. In this point, we have to mention that these results might be different if our samples were larger or different in age. However, it is believed that users' new generation is computer educated and capable to use a lot of different software; therefore the ease of use might not be a strong determinant of Perceived Playfulness.

However, previous cross-cultural studies showed that acceptance models such as TAM are influenced by culture. Thus, despite the similarities of the two groups (students of identical course and age), the results support the same idea regarding CBAAM. Figure 3 presents the values of Greece and Mexico regarding the five cultural dimensions of Hofstede (Power distance, Individualism, Masculinity Uncertainty avoidance). We used these values to suggest some ideas and explanations regarding our results delivered by the differences between Greek and Mexican students. Specifically: (1) regarding the items only one item of CSE was problematic. This sustains that the questionnaire's items were reliable and valid for measuring the perceptions of Greek and Mexican students regarding our research

constructs. (2) Social Influence determines Perceived Usefulness only in Mexico. This could be supported from Hofstede's findings regarding Power Distance for Greece equals to 60 and for Mexico equals to 81 (Figure 3). The higher Power Distance for Mexico indicates that Mexicans are individuals that are influenced from their superiors' opinions; therefore social influence effect is larger and significant in Mexico. In our research, this means that students' opinions are influenced by their tutors or other academic staff regarding the use of CBA, while Greek students are not significantly affected by their professors or other staff. (3) Goal Expectancy has a significant positive effect on Perceived Usefulness and on Perceived Playfulness only in Greece. Goal Expectancy is a variable related with someone's effort to be properly prepared regarding CBA. It is believed that students who avoid uncertainty will also be properly prepared. Uncertainty Avoidance is another cultural dimension (Hofstede, 2001) which is higher in Greece and consequently Goal Expectancy has significant effects in Greece (Figure 3). This means that Greek students with expectations regarding the CBA would be more likely to find it useful and playful. On the other hand, Mexican students did not associate their goal expectations with the usefulness and the playfulness of the system. (4) Perceived Ease of Use defines Perceived Usefulness and Behavioral Intention only in Greece. This result is in line with Srite (2006) results which used Hofstede's (2001) findings to support that cultures that are less masculine might be more concerned with the ease of use of a technology. Hofstede's (2001) findings regarding masculinity for Greece equals with 57 and for Mexico equals with 69, therefore the direct positive effect of Perceived Ease of Use on Behavioral Intention is significant only in Greece. The significant effect of Greek students' perceptions regarding Ease of Use on Behavioral Intention to use a CBA and on Perceived Usefulness is in line with previous studies (Terzis & Economides, 2011). (5) Perceived Usefulness explains Behavioral intention only in Mexico. This result could also be explained by Masculinity. In a society with high masculinity such as Mexico, people put more emphasis on goals and tasks, therefore perceived usefulness, a variable that is linked with CBA's capability to reach user's goals and tasks, is an important determinant of behavioral intention to use a CBA in Mexico. The possible cultural effects on relationships among CBAAM's variables are displayed in Figure 4.

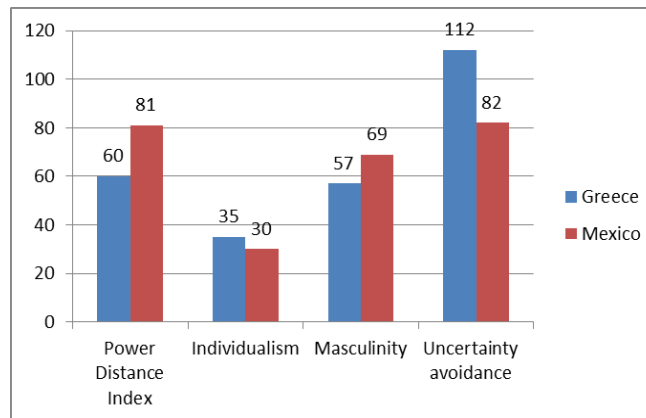


Figure 3. Hofstede's values regarding Greece and Mexico (Hofstede, 2001)

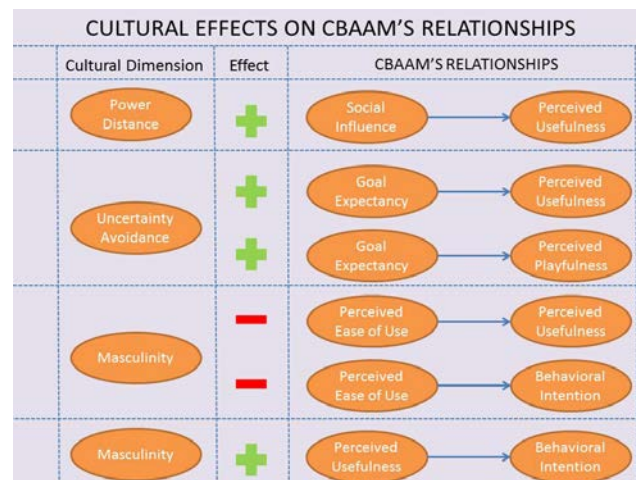


Figure 4. Cultural effects on relationships among CBAAM's variables

## Conclusions

This study compares the user's acceptance behavior of a computer based assessment system in two different cultural environments (Greece and Mexico) by applying the CBAAM (Terzis & Economides, 2011) in both cultures. Despite the good model fit in both countries and the aforementioned similarities which indicate a trend to a globalized use of CBA systems, ethnic or national culture plays important role on user's behavioral intentions regarding CBA acceptance.

This research faces some limitations which might have influenced the results. The first limitation is the small number of individuals regarding Mexico's sample. A sample with more students might have provided different and more significant results. Moreover, the sample is very specific. All the participants are first-year undergraduate students in an introductory course to informatics. Similar studies should be applied to other groups with different characteristics regarding age, specialization, nationality and course's content. Thus, the results should be treated as indications and not as proofs.

To conclude, this study provides potential evidences regarding the cultural effect on CBA acceptance. Developers, researchers and educators could found helpful the results and they should take them into consideration for future (1) CBA systems' development, (2) studies regarding the implementation and acceptance of LMS and CBA systems, especially studies interest in cultural dimensions and their effects in education, (3) the implementation of new and more personalized educational practices. Further studies should be towards this direction in order to provide more useful and significant results.

## References

- Agarwal R., Sambamurthy V. & Stair R.M. (2000). Research report: The evolving relationship between general and specific computer self-efficacy—an empirical assessment. *Information Systems Research*, 11(4), 418–430.
- Ajzen, I. (1991). The theory of planned behaviour. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211.
- Al-Gahtani, S. S., Hubona G. S., & Wang J. (2007). Information technology (IT) in Saudi Arabia: Culture and the acceptance and use of IT. *Information and Management*, 44(8), 681–691.
- Bandyopadhyay, K., & Bandyopadhyay, S. (2008, March). *User acceptance of information technology across cultures*. Paper Presented at the South West Decision Sciences Institute (SWDSI), Houston, USA. Retrieved from <http://www.swdsi.org/swdsi08/paper/AuthorList.htm>
- Barclay, D., Higgins, C., & Thompson, R. (1995). The partial least squares approach to causal modelling: Personal computer adoption and use as an illustration. *Technology Studies*, 2(1), 285–309.
- Chin, W. W. (1998). The partial least squares approach to structural equation modelling. In G. A. Marcoulides (Ed.), *Modern Business research Methods* (pp. 295–336). Mahwah, NJ: Lawrence Erlbaum Associates.
- Cohen, J. (1988). *Statistical power analysis for the behavioural Sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Compeau, D. R., & Higgins, C. A. (1995). Computer self-efficacy: Development of a Measure and Initial Test. *MIS Quarterly*, 19(2), 189–211.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340.
- Falk, R. F., & Miller, N. B. (1992). *A primer for soft modeling*. Akron, OH: University of Akron Press.
- Fishbein, M., & Ajzen, I. (1975). *Belief, Attitude, intention and Behavior: An Introduction to theory and research*. Reading, MA: Addison-Wesley.
- Frankola, K. (2000). *Why online learners drop out*. Retrieved June 07, 2011, from [http://www.kfrankola.com/Documents/Why%20online%20learners%20drop%20out\\_Workforce.pdf](http://www.kfrankola.com/Documents/Why%20online%20learners%20drop%20out_Workforce.pdf)
- Gaspay, A., Dardan S., & Legorreta L. (2007). "Software of the mind" – A review of applications if Hofstede's theory to IT research. *Journal of Information Technology Theory and Application*, 9(3), 1–37.
- Gvozdenko, E., & Chambers, D. (2007). Beyond test accuracy: Benefits of measuring response time in computerised testing. *Australasian Journal of Educational Technology*, 23(4), 542–558.

- Hall, E. T. (1989). *Beyond Culture*. New York, NY: Anchor Books Editions.
- Hofstede, G., & Bond, M. H. (1988). The Confucius connection: From cultural roots to economic growth. *Organizational Dynamics*, 16(4), 4-21.
- Hofstede, G. (2001). *Culture's Consequences: Comparing values, behaviors, institutions, and organizations across nations - 2nd Edition*. Thousand Oaks, CA: Sage Publications.
- Huang, L., Lu M.-T., & Wong, B. K. (2003). *Testing of the cross-cultural applicability of technology acceptance model: Evidence from the PRC, information technology and organizations*. Hershey, New Jersey, USA, Idea Group Publishing.
- Im, I. Hong, S., & Kang, M. S. (2011). An international comparison of technology adoption, testing the UTUAT model. *Information and Management*, 48(1), 1-8.
- Joosten-ten Brinke, D., van Bruggen, J., Hermans, H., Burgers, J., Giesbers, B., Koper, R., & Latour, I. (2007). Modeling assessment for re-use of traditional and new types of assessment. *Computers in Human Behavior*, 23(6), 2721-2741.
- Karahanna E., & Straub D.W., (1999). The psychological origins of perceived usefulness and ease of use. *Information and Management*, 35, 237-50.
- Kaklauskas, A., Zavadskas, E. K., Pruskus, V., Vlasenko, A., Seniut, M., & Kaklauskas, G. (2010). Biometric and intelligent self-assessment of student progress system. *Computers & Education*, 55(2), 821-833.
- Keller, C., Hrastinski, S., & Carlsson, S.A. (2007). Students' acceptance of e-learning environments: A comparative study in Sweden and Lithuania. *Proceedings of the 15th European Conference on Information Systems* (pp. 395-406). Retrieved from <http://aisel.aisnet.org/ecis2007/40>
- Landry, B. J. L., Griffeth, R., & Hartman, S. (2006). Measuring student perceptions of blackboard using the technology acceptance model. *Decision Sciences Journal of Innovative Education*, 4(1), 87-99.
- Lee, Y. C. (2008). The role of perceived resources in online learning adoption. *Computers & Education*, 50(4), 1423-1438.
- Li, X., Hess, T. J., McNab, A. L., & Yu, Y. (2009). Culture and acceptance of global web sites: A cross-country study of the effects of national cultural values on acceptance of a personal web portal. *SIGMIS Database*, 40(4), 49-74.
- Moon, J., & Kim, Y. (2001). Extending the TAM for a world-wide-web context. *Information and Management*, 38(4), 217-230.
- Moridis, C. N., & Economides, A. A. (2009b). Prediction of student's mood during an online test using formula-based and neural network-based method. *Computers & Education*, 53(3), 644-652.
- Ong, C., & Lai, J. (2006). Gender differences in perceptions and relationships among dominants of e-learning acceptance. *Computers in Human Behaviour*, 22(5), 816-829.
- Ong, C.-S. Lai, J.-Y., & Wang, Y.-S. (2004). Factors affecting engineers' acceptance of asynchronous e-learning systems in high-tech companies. *Information and Management*, 41(6), 795-804.
- Padilla-Melendez, A., Garrido-Moreno, A., & Del Aguila-Obra, A. R. (2008). Factors affecting e-collaboration technology use among management students. *Computers & Education*, 51(2), 609-623.
- Ringle, C. M., Wende, S., & Will, A. (2005). *SmartPLS 2.0 (beta)*. Retrieved from <http://www.smartpls.de>
- Shee, D. Y., & Wang, Y.-S. (2008). Multi-criteria evaluation of the web-based e-learning system: A methodology based on learner satisfaction and its applications. *Computer & Education*, 50(3), 894-905.
- Singh, N., Fassott, G., Chao, M. C. H., & Hoffmann, J. A. (2006). Understanding international Website usage: A cross-national study of German, Brazilian, and Taiwanese online consumers. *International Marketing Review*, 23(1), 83-97.
- Smith, B., & Caputi, P. (2007). Cognitive interference model of computer anxiety: Implications for computer based assessment. *Computers in Human Behavior*, 23(3), 713-728.
- Srite, M., & Karahanna, E. (2006). The role of espoused national cultural values in technology acceptance. *MIS Quarterly*, 30(3), 679-704.
- Straub, D., Keil, M., & Brenner, W. (1997). Testing the technology acceptance model across cultures: A three country study. *Information and Management*, 33(1), 1-11.
- Straub, D.W., Loch, K., Evaristo, R., Karahanna, E., & Srite, M. (2002). Toward a theory-based measurement of culture. *Journal of Global Information Management*, 10(1), 13-23.



- Taksa, I., & Flomenbaum, J. M. (2009). An integrated framework for research on cross-cultural information retrieval. *Proceedings of the Sixth International Conference on Information Technology: New Generations* (pp. 1367-1372). doi: 10.1109/ITNG.2009.99
- Taylor, S., & Todd, P. A. (1995). Understanding information technology usage: A test of competing models. *Information Systems Research*, 6(2), 144–176.
- Tenenhaus, M., Vinzi, V. E., Chatelin, Y.-M., & Lauro, C. (2005). PLS path modeling. *Computational Statistics and Data Analysis*, 48(1), 159–205.
- Teo, T. (2009). Modelling technology acceptance in education: A study of pre-service teachers. *Computers & Education*, 52(1), 302–312.
- Teo, T., Lee, C. B., & Chai, C. S. (2008). Understanding pre-service teachers' computer attitudes: Applying and extending the technology acceptance model. *Journal of Computer Assisted Learning*, 24(2), 128–143.
- Teo, T., Luan, W. S., & Sing, C.C. (2008). A cross-cultural examination of the intention to use technology between Singaporean and Malaysian pre-service teachers: An application of the Technology Acceptance Model (TAM). *Educational Technology and Society*, 11(4), 265–280.
- Terzis, V., & Economides, A. A. (2011). The acceptance and use of computer based assessment. *Computers & Education*, 56(4), 1032–1044.
- Trompenaars, F., & Hampden-Turner, C. (1988). *Riding the waves of culture*. New York, NY: McGraw-Hill.
- Van Raaij, E. M., & Schepers, J. J. L. (2008). The acceptance and use of a virtual learning environment in China. *Computers & Education*, 50(3), 838–852.
- Venkatesh, V. (1999). Creation of favorable user perceptions: Exploring the role of intrinsic motivation. *MIS Quarterly*, 23, 239–260.
- Venkatesh, V., & Davis F. D. (1996). A model of the antecedents of perceived ease of use: Development and test. *Decision Sciences*, 27, 451–481.
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478.
- Wang, Y. (2003). Assessment of learner satisfaction with asynchronous electronic learning systems. *Information & Management*, 41(1), 75–86.
- Wang, Y.-S., Wu, M.-C., & Wang, H.-Y. (2009). Investigating the determinants and age and gender differences in the acceptance of mobile learning. *British Journal of Educational Technology*, 40(1), 92–118.
- Wetzels, M., Odekerken-Schröder, G., & Van Oppen, C. (2009). Using PLS path modelling for assessing hierarchical construct models: Guidelines and empirical illustration. *MIS Quarterly*, 33(1), 177–195.
- Wixon, B. H., & Watson, H. J. (2001). An empirical investigation of the factors affecting data warehousing success. *MIS Quarterly*, 25(1), 17–41.
- Wold, H. (1982). Soft Modeling: The basic design and some extensions. In Karl G. Jöreskog & Herman Wold (Eds.), *Systems under indirect observation: Causality, structure prediction II* (pp. 1–54). Amsterdam, Netherlands: North Holland.
- Yi, M. Y., & Hwang, Y. (2003). Predicting the use of web-based information systems: Self-efficacy, enjoyment, learning goal orientation, and the technology adoption model. *International Journal of Human Computer Studies*, 59(4), 431–449.
- Zakour, & Amel B. (2004). Cultural differences and information technology acceptance. *Proceedings of the 7th Annual Conference of the Southern Association for Information Systems* (pp. 156-161). Retrieved from <http://aisel.aisnet.org/sais2004/26>.

## Appendix 1

Constructs	Items	
Perceived Usefulness	PU1	Using the Computer Based Assessment (CBA) will improve my work.
	PU2	Using the Computer Based Assessment (CBA) will enhance my effectiveness.
	PU3	Using the Computer Based Assessment (CBA) will increase my productivity.
Perceived Ease of Use	PEOU1	My interaction with the system is clear and understandable.
	PEOU2	It is easy for me to become skilful in using the system.
	PEOU3	I find the system easy to use.
Computer Self Efficacy	CSE1	I could complete a job or task using the computer.
	CSE2	I could complete a job or task using the computer if someone showed how to do it first.
	CSE3	I can navigate easily through the Web to find any information I need.
	CSE4	I was fully able to use the computer and Internet before I began using the Computer Based Assessment (CBA).
Social Influence	SI1	People who influence my behaviour think that I should use CBA.
	SI2	People who are important to me think that I should use CBA.
	SI3	The seniors in my university have been helpful in the use of CBA.
	SI4	In general, my university has supported the use of CBA.
Facilitating Conditions	FC1	When I need help to use the CBA, someone is there to help me.
	FC2	When I need help to learn to use the CBA, system's help support is there to teach me.
Content	C1	CBA's questions were clear and understandable.
	C2	CBA's questions were easy to answer.
	C3	CBA's questions were relative to the course's syllabus.
	C4	CBA's questions were useful for my course.
Goal Expectancy	GE1	My Course's preparation was sufficient for the CBA
	GE2	My personal preparation for the CBA.
	GE3	My performance expectations for the CBA.
Perceived Playfulness	PP1	Using CBA keeps me happy for my task.
	PP2	Using CBA gives me enjoyment for my learning.
	PP3	Using CBA, my curiosity is stimulated.
	PP4	Using CBA will lead to my exploration.
Behavioural Intention to use CBA	BI1	I intend to use CBA in the future.
	BI2	I predict I would use CBA in the future.
	BI3	I plan to use CBA in the future.

## Perspectives on Open and Distance Learning: Open Educational Resources: An Asian Perspective

(Book Review)

### **Reviewer:**

**Krista Jensen**

Knowledge Mobilization Officer, York University

PhD Student, Faculty of Education, York University

krista\_jensen@edu.yorku.ca

### **Textbook Details:**

Perspectives on Open and Distance Learning: Open Educational Resources: An Asian Perspective

Edited by Gajaraj Dhanarajan and David Porter

2013, 274 pages, Commonwealth of Learning, Vancouver, Canada, [www.col.org](http://www.col.org) and OER Asia, Penang, Malaysia, [www.oerasia.org](http://www.oerasia.org)

ISBN: 978-1894975612 (PDF) <http://www.col.org/resources/publications/Pages/detail.aspx?PID=441>

'Perspectives on Open and Distance Learning: Open Educational Resources: An Asian Perspective' offers a look at the current utilization of open educational resources (OER) in higher education (HE) in Asia. This compilation of country perspectives and cases studies is co-published by the Commonwealth of Learning (COL) in Vancouver, Canada and OER Asia operated out of the Institute for Research and Innovation at the Wawasan Open University in Malaysia. It is the latest in a series of publications addressing OER from the COL, which is an intergovernmental organisation comprised of more than 50 countries whose mandate is to "encourage the development and sharing of open learning/distance education knowledge, resources and technologies" ([www.col.org/about](http://www.col.org/about)).

While there are several different definitions of what exactly OER are, the common thread among all the proposed explanations is they are educational resources, from individual lesson plans to entire course modules, that are openly available and can be used by teachers and students for free. Further to this, OER users are actively encouraged to reuse, revise, remix and redistribute the resources, although in reality there are often barriers to accomplishing this goal.

Following a brief overview of the state of OER in HE in Asia generally, the book includes ten in-depth country perspectives, in addition to ten case studies showcasing specific uses of OER. The perspectives and cases studies highlight both the opportunities and barriers of using OER in HE. While some institutions have been quicker to adopt the use of OER than others, it is evident that the use of these resources is on the rise and major investments are being made to integrate them into the HE environment in Asia.

The country perspectives provide a detailed report on the use of OER in various HE institutions in China, Hong Kong, India, Indonesia, Japan, Korea, Malaysia, Pakistan, Philippines and Vietnam. Every chapter in this section presents an overview of how OER have been adopted in each of the ten countries, followed by a survey conducted by the authors to further explore the level of acceptance and engagement of OER by HE institutions, and more especially by faculty members and teachers. While the results vary from country to country and from institution to institution, a number of issues appear consistently.

Copyright and intellectual property concerns are cited by all the authors, with the exception of the Virtual University of Pakistan, which makes all of its OER courses available under a Creative Commons license and requires OER producers to assign all intellectual property rights over to the university. In many of the countries perspectives, the authors identify a need for education on copyright and intellectual property rights for faculty members. The surveys indicate that many do not fully understand the implications of both using OER or producing and contributing OER. And while almost all of the HE institutions reported upon are either already using Creative Commons licenses or in the process of moving in that direction, there is also a lack of understanding of what this actually means for individual faculty members. I believe part of the challenge is this is a major shift in the way most faculty and HE institutions are used to working; a shift away from having personal ownership and control over educational resources to making those resources openly available with little to no control over how they are used.

Quality assurance is also cited as an ongoing concern and potential barrier to making educational resources openly available. Understandably, HE institutions and individuals producing OER want the materials they make available to be of excellent quality. Those who succeed in producing high quality materials do so with considerable coordination, time and financial resources. The University of Madras provides a valuable example of a project that entailed producing 250 openly available instructional modules related to students' soft skills. The project had a budget of \$70,320USD and took 27 months to complete with three working groups, including project personnel, a core content development team and an advisory committee to provide quality assurance checks. The author reports that at the time of writing this report, no major cost benefits existed, mainly due to the expense of producing the large number of OER. It seems it is simply too early to tell what the cost benefits are of producing OER on an institution wide scale, however response from teachers, students and the university administration has been positive.

Encouraging faculty and teachers to contribute OER material can be difficult. From the surveys conducted, it appears that while faculty members are actively using OER in their classrooms, they are not yet in the habit of making their teaching materials freely available. Several authors mention the need for HE to create institutional policies that support and promote not only the use of OER, but also the development of such resources amongst their faculty members. Related to this is the need for HE to offer supports to make this culture shift easier, by way of offering training, tech support and scholarly incentives.

Another way to help ease this culture shift would be to provide more knowledge sharing opportunities for faculty and teachers using and producing OER. Many authors reported a lack of organized ways for faculty to learn from each other, both within individual institutions and across institutions and countries. Using readily available social media tools to connect people and provide them with a chance to share their experiences, as well as pose and answer questions may help spread the use of OER more widely and allow faculty to become more comfortable with using them.

The second section of the book offers ten case studies from Malaysia, the Philippines, Korea, Indonesia, Hong Kong, China and four different institutions in India. These case studies provide a glimpse into how OER are being adopted and adapted to suit local needs. Each location has developed different approaches making OER openly available. Many have their institutionally developed OER on purpose built systems, for example Korea's Open Courseware System and the Indira Gandhi National Open University's eGyanKosh, a learning resource repository; while others have chosen to instead use third party solutions such as YouTube and iTunesU, as is the case with Beijing Open University. What emerges from these case studies however, is that the exact technology used is really a matter of what makes the most sense for each location—this isn't a one-size-fits-all endeavour. More important is that institutions are able to make OER freely available in a way that builds on their own individual needs and strengths. In the countries that are most successfully embracing OER, as demonstrated in India, Pakistan and Malaysia, their use has partially been driven by a desire to provide access to education for all members of society, especially to those who have previously been excluded from access to HE. Emerging markets need to provide training and education to a wider audience in new ways in order to achieve their desire for economic growth and equality. This desire, coupled with supportive policies and major financial investments from HE, government sectors and other partners, is enabling countries like those mentioned above to emerge as leaders in OER in Asia.

One of the main strengths of this book is the wide range of examples presented from across Asia, representing a number of different cultures, educational systems and economic realities. I thought the case studies were particularly useful, as they feature very pragmatic issues such as how a variety of OER programs or systems were implemented, how they were supported (or not), challenges faced, lessons learned and the opportunities and successes that were identified, all of which would be of interest to those working in this area.

Unfortunately some of the surveys are more robust than others and some contain relatively low numbers of respondents, for example just 15 respondents are identified in the Japanese survey. This makes it difficult to determine just how wide spread the use OER really is in HE. It would have also been useful to survey students at the various institutions to get a sense of how they are using the OER and how important they feel these resources are to their success. I would have also liked to see case studies from all the countries surveyed in the country perspectives section, as there are no case studies from Japan, Pakistan or Vietnam. And lastly, given the rich insights and learnings presented in the case studies, a list of best practices or issues to be mindful of when initiating an OER project would have been a valuable inclusion.

'Perspectives on Open and Distance Learning: Open Educational Resources: An Asian Perspective' provides a valuable overview of the current state of OER use in Asia, in addition to highlighting common issues, barriers and opportunities. While the book focuses on Asian examples, I believe the points raised would be of use those in the HE field worldwide. The country perspectives and case studies are especially useful in demonstrating the value of investing in this emerging area of educational technology, as well as the benefits of providing open and freely available OER to populations who might not otherwise have access to HE.

This book would be useful to faculty and teachers, as well as not-for-profits, NGOs and HE institutions interested in using and developing OER. It would also provide useful insights to policy makers, both at universities and in government who can mandate and fund the use of OER. Researchers interested in educational technology and OER, especially those looking for groups or environments to study, may also find this resource of value.